

UNIVERSITY OF ALBERTA

High Data Rate Transmission in Multi-Service Code Division Multiple
Access Wireless Communication Systems

By

Sun, Songsong



A thesis submitted to the Faculty of Graduate Studies and Research in partial
fulfillment of the requirements for the degree of Doctor of Philosophy

Department of Electrical and Computer Engineering

Edmonton, Alberta

Spring, 1999



National Library
of Canada

Acquisitions and
Bibliographic Services

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque nationale
du Canada

Acquisitions et
services bibliographiques

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence

Our file Notre référence

The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/film, de reproduction sur papier ou sur format électronique.

L'auteur conserve la propriété du droit d'auteur qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

0-612-39596-0

Canada

University of Alberta

Library Release Form

Name of Author: Sun, Songsong

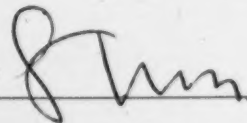
Title of Thesis: High Data Rate Transmission in Multi-Service Code Division Multiple
Access Wireless Communication Systems

Degree: Doctor of Philosophy

Year this Degree Granted: 1999

Permission is hereby granted to the University of Alberta Library to reproduce single copies of this thesis and to lend or sell such copies for private, scholarly, or scientific research purposes only.

The author reserves all other publication and other rights in association with the copyright in the thesis, and except as hereinbefore provided, neither the thesis nor any substantial portion thereof may be printed or otherwise reproduced in any material form whatever without the author's prior written permission.



Sun, Songsong

710, Ru Yi Li No. 8

Nanjing, China 210018

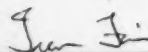
Date:

Jan. 28, 1999

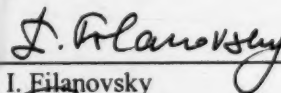
UNIVERSITY OF ALBERTA

FACULTY OF GRADUATE STUDIES AND RESEARCH

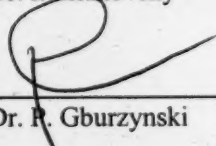
The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research for acceptance, a thesis entitled **High Data Rate Transmission in Multi-Service Code Division Multiple Access Wireless Communication Systems** submitted by Songsong Sun in partial fulfillment of the requirements for the degree of **Doctor of Philosophy**.



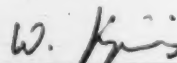
Dr. I. Fair



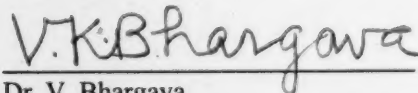
Dr. I. Eilanovsky



Dr. R. Gburzynski



Dr. W. Krzymien
Supervisor



Dr. V. Bhargava
External Examiner

Date: 28 Jan. 1999

In memory of my father
Sun, Shou Qin

献给我的父亲孙守勤

ABSTRACT

The first and second generation wireless systems are mainly designed for voice service. With the expanding usage of Internet, the demand for wireless data service has been the driving force for the design and development of third generation wireless systems, which are required to support voice, as well as circuit and packet data services. The code division multiple access (CDMA) technique has been accepted as the multiple access scheme for the third generation wireless standards all over the world. This Ph.D. thesis addresses the issues of high data rate transmission in multi-service CDMA wireless communication systems. The study involves the design and analysis of efficient access schemes, fast chip synchronizers, radio link protocols and call admission control policies, system capacity analysis, optimal transmitted power allocation, as well as the optimization of the signal-to-interference ratio assignment.

ACKNOWLEDGEMENT

First and foremost, I thank my supervisor Dr. Witold Krzymien for providing me with an exceptional environment to develop my professional skills and to grow intellectually. The financial support from Dr. Krzymien, *TRLabs*, and the University of Alberta is also much appreciated. This work could never have been done without their support.

I also thank Peng Mok for providing me with some of the simulation source codes. Without his generous help, this work would not have been done so quickly.

I wish to thank Dr. Vijay Bhargava for his excellent comments, which help improve the quality of this thesis. Dr. Ahmad Jalali is also acknowledged for his advice and many stimulating discussions.

Finally, I wish to thank all the staff and students at *TRLabs* for making the *TRLabs* experience truly enjoyable and memorable.

Table of Contents

Chapter 1	INTRODUCTION	1
1.1	Historical review	1
1.2	Wireless systems and standards	1
1.2.1	Introduction to multiple access schemes	2
1.2.1.1	Frequency division multiple access (FDMA)	2
1.2.1.2	Time division multiple access (TDMA)	4
1.2.1.3	Code division multiple access (CDMA)	4
1.2.2	First generation cellular systems	5
1.2.3	Second generation cellular systems	7
1.2.3.1	North-American TDMA digital cellular (IS-54,136)	7
1.2.3.2	Global system for mobile communications (GSM)	8
1.2.3.3	CDMA digital cellular standard (IS-95)	8
1.2.4	Third generation cellular standard	10
1.3	Thesis objectives	13
1.4	Thesis contributions	16
	References	18
Chapter 2	DESIGN AND PERFORMANCE EVALUATION OF ACCESS SCHEMES	21
2.1	Introduction	21
2.2	Proposed access protocols.	24
2.2.1	SDTX-MAC protocol	24
2.2.2	ADTX-MAC protocol	26

2.2.3	HDTX-MAC protocol	27
2.3	Structure of access request messages	27
2.4	Synchronizer circuits	28
2.4.1	Background	28
2.4.2	Modified synchronizer using matched filter techniques	36
2.4.3	Performance analysis	40
2.5	Delay/throughput analysis	48
2.5.1	SDTX-MAC scheme	48
2.5.2	ADTX-MAC scheme	51
2.5.3	HDTX-MAC Scheme	61
2.6	Performance evaluation	62
2.7	Conclusions and discussion	67
	References	68
Chapter 3	DATA TRANSMISSION WITH HIGH RELIABILITY	72
3.1	Introduction	72
3.2	System description	74
3.2.1	Physical layer description	74
3.2.2	Retransmission protocols	80
3.2.2.1	Basic automatic-repeat-request (ARQ) strategies	80
3.2.2.2	Hybrid ARQ	84
3.2.2.3	Retransmission schemes	86
3.3	Performance analysis	88
3.3.1	Error detection	88

3.3.2	Error correction	94
3.3.3	Power control	100
3.3.4	Throughput and reliability	102
3.4	Optimum signal-to-interference ratio assignment	105
3.5	Simulation study	108
3.6	Conclusions	116
	References	117
Chapter 4	CAPACITY ANALYSIS AND CALL ADMISSION CONTROL	122
4.1	Introduction	122
4.2	Capacity analysis	125
4.2.1	Capacity analysis for conventional multiple access systems	125
4.2.2	Capacity analysis for entirely interference limited systems	129
4.2.2.1	Intuitive CDMA capacity analysis	129
4.2.2.2	Detailed capacity analysis	133
4.3	Call admission control (CAC) policies	140
4.3.1	SIR-based CAC algorithm	141
4.3.2	General call admission control (CAC) criterion	143
4.3.3	Non-predictive CAC (NPCAC)	147
4.3.4	Predictive CAC (PCAC)	148
4.3.5	Simulation and numerical results	150
4.4	Conclusions	156

	References	156
Chapter 5	OPTIMAL FORWARD LINK POWER ALLOCATION	159
5.1	Introduction	159
5.2	Forward link system description	161
5.3	Optimal power allocation	166
5.3.1	Problem formulation	166
5.3.2	Numerical results	168
5.4	Conclusion and discussion	175
	References	177
Chapter 6	CONCLUSIONS AND FUTURE RESEARCH DIRECTIONS	178
	References	181

List of Tables

Table 1.1. Wideband IS-95 parameters	12
Table 1.2. European/Japanese wideband CDMA parameters	13
Table 3.1. Weight distribution of the dual code of the BCH (255, 231) binary code	92
Table 3.2. Probability of undetected error (P_{ue})	93
Table 3.3. Parameters and optimum SIRs	116
Table 4.1. Parameters used to obtain the numerical results of Erlang capacities	137
Table 4.2. Parameters used in the simulation and numerical results for CAC algorithms	151
Table 4.3. Bandwidth available for home and remote cells corresponding to different values of ρ	153
Table 5.1. Parameters used to obtain numerical results for the optimal power allocation	171
Table 5.2. Numerical results when the criterion is to minimize the mean transmit power	175
Table 5.3. Numerical results when the criterion is to minimize the sum of the mean transmit power and its standard deviation	175

List of Figures

Fig. 1.1. Frequency division multiple access.	3
Fig. 1.2. Time division multiple access.	3
Fig. 1.3. Code division multiple access.	5
Fig. 1.4. Illustration of the cellular frequency reuse concept.	6
Fig. 1.5. Illustration of multicarrier and direct spread transmissions	11
Fig. 2.1. Traffic characteristics of voice and data.	21
Fig. 2.2. Structure and parameters of the SDTX-MAC protocol.	25
Fig. 2.3. Structure and parameters of the ADTX-MAC protocol.	26
Fig. 2.4. Structure and parameters of the HDTX-MAC protocol.	27
Fig. 2.5. Time or phase uncertainty region.	29
Fig. 2.6. Block diagram of the single dwell time PN acquisition system.	30
Fig. 2.7. Block diagram of the multiple dwell time PN acquisition system.	31
Fig. 2.8. A matched filter acquisition synchronizer.	33
Fig. 2.9. Discrete-time implementation of a filter matched to an N -chip spreading waveform.	35
Fig. 2.10. Oversampled discrete-time matched filter.	35
Fig. 2.11. BPSK direct sequence spread spectrum transmitter.	37
Fig. 2.12. QPSK direct sequence spread spectrum transmitter.	37
Fig. 2.13. Block diagram of the synchronizer for BPSK spreading.	38
Fig. 2.14. Block diagram of the synchronizer for QPSK spreading.	38
Fig. 2.15. Differential discrete-time matched filter.	39
Fig. 2.16. Simplified interference model.	42

Fig. 2.17. Probability of correct acquisition in the Gaussian channel.	48
Fig. 2.18. Probability of correct acquisition in the Rayleigh fading channel.	48
Fig. 2.19. Packets from two users on a pure ALOHA channel.	53
Fig. 2.20. Packets from two users on a slotted ALOHA channel.	54
Fig. 2.21. Throughput versus offered traffic for the pure and slotted ALOHA.	55
Fig. 2.22. Autocorrelation function for a maximum-length PN sequence.	55
Fig. 2.23. Throughput versus offered traffic for a spread ALOHA channel.	58
Fig. 2.24. Mean access delay versus the initial ARM's SNR for the SDTX-MAC protocol.	64
Fig. 2.25. Mean access delay versus the initial ARM's SNR for the ADTX-MAC.	65
Fig. 2.26. Mean access delay versus the initial ARM's SNR for ADTX-MAC protocol.	65
Fig. 2.27. Throughput efficiency versus the initial ARM's SNR for SDTX-MAC.	66
Fig. 2.28. Throughput versus offered traffic for the ADTX-MAC protocol.	66
Fig. 3.1. Block diagram of a DS-CDMA data transmission system.	75
Fig. 3.2. Quadrature spreading of DS-CDMA signals.	76
Fig. 3.3. Block diagram of the RAKE receiver.	77
Fig. 3.4. The l -th branch of the RAKE receiver for M -ary orthogonal signaling.	78
Fig. 3.5. SW-ARQ.	81
Fig. 3.6. GBN-ARQ with $N=4$.	82
Fig. 3.7. SR-ARQ.	84
Fig. 3.8. Example of the operation of the NACK protocol.	88
Fig. 3.9. Constraint length 9, rate $1/3$ convolutional encoder.	95

Fig. 3.10. A constraint length of 3, rate 1/3 convolutional encoder.	95
Fig. 3.11. Trellis diagram for the encoder in Fig. 3.10.	96
Fig. 3.12. Block diagram of a communication system using a convolutional code.	97
Fig. 3.13. Probability of a packet being declared undeliverable (P_{ud}) versus packet error rate (P_F).	105
Fig. 3.14. Throughput efficiency versus SIR.	111
Fig. 3.15. BER (p) versus SIR for the 64-ary orthogonal signaling with rate 1/3 convolutional coding.	112
Fig. 3.16. FER (P_F) versus SIR for the 64-ary orthogonal signaling with rate 1/3 convolutional coding.	113
Fig. 3.17. Accepted packet error rate $P(E)$ versus SIR.	113
Fig. 3.18. Comparison of throughput efficiency vs. SIR for Doppler shift of 5 Hz.	114
Fig. 3.19. Comparison of throughput efficiency vs. SIR for Doppler shift of 15 Hz.	114
Fig. 3.20. Average cumulative signal-to-interference ratio versus SIR per transmission.	115
Fig. 4.1. State transition diagram for the $M/M/c/c$ system.	126
Fig. 4.2. State transition diagram for the $M/M/c$ system.	128
Fig. 4.3. Erlang capacities for voice users, short packet data, and bursty ($p_l = 0.2$) long packet data users.	138
Fig. 4.4. Erlang capacities for voice users, short packet data, and regular ($p_l = 0.4$) long packet data users.	139

Fig. 4.5. Illustration of the SIR-based CAC algorithm.	143
Fig. 4.6. Cell layout.	146
Fig. 4.7. Samples of actual and predicted SIR.	152
Fig. 4.8. Impact of the ARQ on the system's performance.	154
Fig. 4.9. Performance comparison of two call admission control schemes.	155
Fig. 5.1. Illustration of power utilization of data users.	160
Fig. 5.2. Block diagram of the forward link CDMA data transmission system.	162
Fig. 5.3. Modulation and spreading for the pilot signal.	163
Fig. 5.4. Coding, modulation and spreading for the traffic signal.	163
Fig. 5.5. The cell layout.	170
Fig. 5.6. Normalized effective average versus normalized link transmitted power for the mobile at location A.	171
Fig. 5.7. Normalized effective average versus normalized link transmitted power for the mobile at location B.	172
Fig. 5.8. Normalized effective versus normalized link transmitted power for the mobile at location C.	173

List of Abbreviations

ACK	Acknowledgement
ADTX	Asynchronous discontinuous transmission
AMPS	Advanced mobile phone service
ARC	Access reservation channel
ARM	Access request message
ARQ	Automatic-repeat-request protocol
AWGN	Additive white Gaussian noise
BER	Bit error rate
BPSK	Binary phase shift keying
BS	Base station
CAC	Call admission control
CDMA	Code division multiple access
CRC	Cyclic redundancy code
CTX	Continuous transmission
DDMF	Differential digital matched filter
$\pi / 4$ DQPSK	$\pi / 4$ shifted, differential encoded quadrature phase shift keying
DS	Direct sequence
DTX	Discontinuous transmission
ESN	Electronic serial number
FDD	Frequency multiplexing division
FDMA	Frequency division multiple access
FEC	Forward error correcting

FER	Frame error rate
FSM	Finite state machine
GBN	Go-back-N
GMSK	Gaussian-filtered minimum shift keying
GSM	Global system for mobile communications
HDTX	Hybrid discontinuous transmission
MAC	Medium access control
MS	Mobile station
NACK	Negative acknowledgement
NPCAC	Non-predictive call admission control
PCAC	Predictive call admission control
PN	Pseudo-noise
QoS	Quality of service
QPSK	Quadrature phase shift keying
RLP	Radio link protocol
SDTX	Synchronous discontinuous transmission
SIR	Signal-to-interference ratio
SNR	Signal-to-noise ratio
SR	Selective-repeat
SW	Stop-and-wait
TCP/IP	Transmission control/Internet protocols
TDD	Time multiplexing division
TDMA	Time division multiple access

Chapter 1

INTRODUCTION

1.1 Historical review

Wireless communication is enjoying rapid expansion and improvements of its infrastructure, and growth of its market penetration. This expansion has been enabled by technological improvements, such as the development of highly reliable solid-state radio frequency hardware and very large scale integrated digital circuits. The origins of wireless communications go back more than a century when Guglielmo Marconi demonstrated radio's ability to provide continuous communications with ships sailing the English channel. AT&T Bell Laboratories and other research centers in the world developed the cellular concept in the 1960s and 1970s [1], [2], [24] and the first U.S cellular telephone system called Advanced Mobile Phone Service (AMPS) [1] was standardized and deployed in late 1983. Since then, the wireless communications industry has grown by orders of magnitude. It is anticipated that in the first couple of decades of the next century, there will be an equal number of wireless and wireline customers throughout the world [24].

1.2 Wireless systems and standards

Many countries deployed first generation cellular systems in the years around 1980. For example, the Nippon Telephone and Telegraph (NTT) analog cellular system became operational in 1979, and AMPS was deployed in 1983. Although the first generation cellular systems have been widely accepted throughout the world, they were not designed

to provide the capacity demand for efficient use of limited available frequency spectrum. Digital cellular systems offer large improvements in capacity and system performance [3], [4], and they will eventually replace the existing analog systems. Before introducing wireless standards, it would be quite helpful to explain multiple access techniques, which are used to allow many users to share simultaneously a finite amount of communication channels.

1.2.1 Introduction to multiple access schemes

The key to all multiple access schemes is that signals sharing a communications channel (where a communication channel could be either a frequency band, a time slot, or a unique code) will not create unmanageable mutual interference corrupting the detection process. Three major multiple access techniques have been used in the wireless communications standards, which are briefly explained below.

1.2.1.1 Frequency division multiple access (FDMA)

FDMA assigns different frequency channels to different users, as illustrated in Fig. 1.1. From the figure it can be seen that each user is assigned a unique frequency channel, and no other user can share that unique frequency band during the period of the call. This is the simplest and oldest multiple access scheme. As long as the RF filters work well, signals in the channel, for example band #1 will introduce little interference to those in band #2. The main interference in the FDMA system is introduced by mobiles using the same frequency band from other cells due to frequency re-use, the so-called co-channel interference.

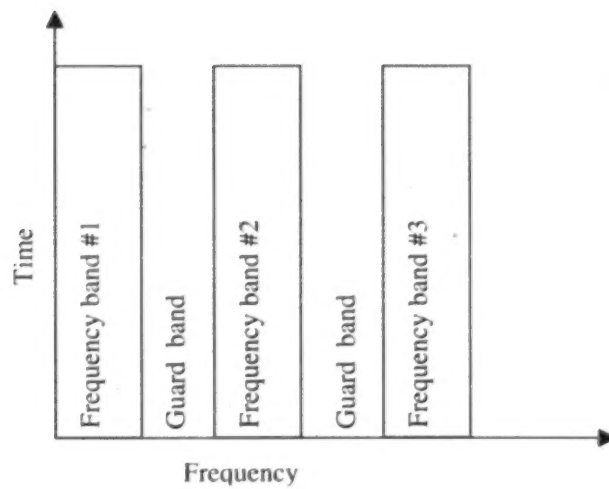


Fig. 1.1. Frequency division multiple access.

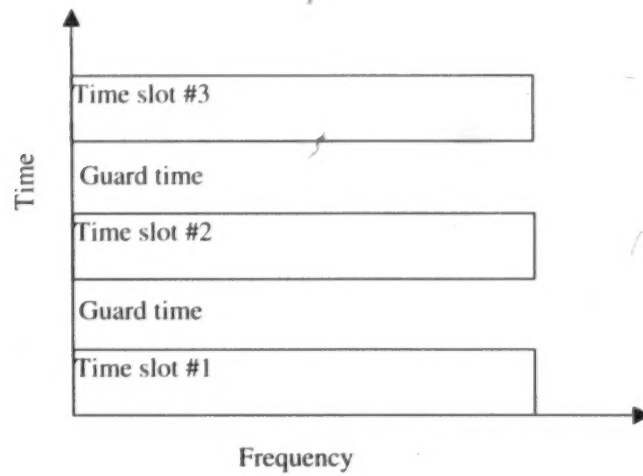


Fig. 1.2. Time division multiple access.

1.2.1.2 Time division multiple access (TDMA)

In TDMA, the communication channel is shared by assigning each user the full frequency band of the communication channel for a short duration of time called a time slot, as shown in Fig. 1.2. A pure TDMA system uses only one carrier frequency. In practice, cellular systems that employ TDMA actually use hybrid FDMA/TDMA. Normally, time axis is segmented into intervals called frames. Each frame is further partitioned into time slots. Each user may periodically employ one or more slots.

1.2.1.3 Code division multiple access (CDMA)

CDMA is a multiple access scheme using spread spectrum modulation techniques [5], [6], which intentionally widen the spectrum of narrowband signals through multiplication by very large bandwidth spreading waveforms. The spreading waveforms have a structure of pseudo random, or pseudo-noise (PN) sequences of very narrow pulses (orders of magnitude narrower than the pulses of the information bearing digital signal), which are known as chips. All user signals in a CDMA system occupy the same band of radio frequencies at the same time. Each mobile user has his own spreading sequence, or his own code which is approximately orthogonal to all other user sequences, or codes. An easy way to explain CDMA is shown in Fig. 1.3. It can be seen from the figure it is the code rather than frequency band or time slot that distinguishes different users. In principle, at the receiving end, the receiver performs time correlation to despread only the specific desired signal. All other user signals remain spread and appear to be noise due to the correlation property of PN codes. This is the so-called direct-sequence (DS) spread spectrum system. There is another type of CDMA system, which is based on frequency-

hopped (FH) spread spectrum. With FH-CDMA the PN sequence is used to pseudo-randomly hop the carrier frequency throughout a large bandwidth.

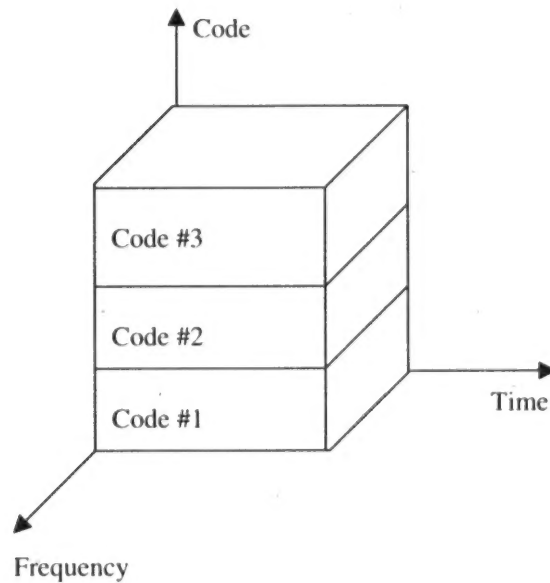


Fig. 1.3. Code division multiple access

1.2.2 First generation cellular systems

First generation cellular systems are based on FDMA and FM technology. The frequency spectrum allocations for these systems are in the 800/900 MHz band. For example, in AMPS, analog voice signals are frequency modulated onto carriers in the 800 MHz band. A mobile station (MS) is actually assigned two channels, one for the forward link (base to mobile) or downlink, and one for the reverse link (mobile to base) or uplink. For example in AMPS, the 824-849 MHz band is allocated for the forward link and the 869-894 MHz band is allowed for the reverse link. A 30 kHz band is allocated to each channel. Each

cell actually uses a subset of the set of RF channels available to the entire system, and neighboring cells must use different sets of channels. This is the concept of cellular frequency reuse, which is illustrated in Fig. 1.4, where cells labeled with the same number use the same group of channels. The frequency reuse factor $1/7$ is used in this case. It is also possible to reuse frequency in a 3-cell, 4-cell or 12-cell cluster. Cell clustering results in reduced cochannel interference, but the system capacity is reduced. In CDMA systems, however, all cells use the same frequency band.

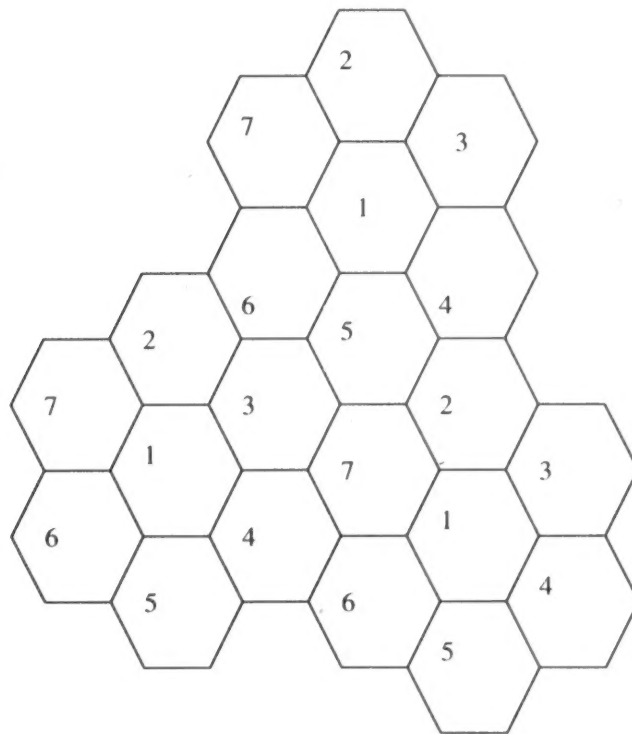


Fig. 1.4. Illustration of the cellular frequency reuse concept.

1.2.3 Second generation cellular systems

Increasing demand for cellular telephone service has pushed analog cellular technology to its capacity limit and resulted in proposals for spectrally efficient second generation cellular systems. Second generation cellular systems use digital modulation and either TDMA/FDMA or CDMA/FDMA techniques. There are three major second generation digital cellular standards. A brief description of these standards follows.

1.2.3.1 North-American TDMA digital cellular standard (IS-54, IS-136)

The IS-54 [7] standard was developed as a replacement for the existing analog AMPS system. The IS-54 channel structure actually uses a combination of FDMA and TDMA. In order to transit smoothly from AMPS to IS-54, the IS-54 system uses the same frequency band and channel spacing as AMPS (carriers spaced 30 kHz apart) and supports as many as six IS-54 users (three users if operating in full data rate mode) on each AMPS channel. In the IS-54 implementation, mobile stations are capable of operating on either AMPS FDMA analog or TDMA/FDMA digital channels. The control channels in IS-54 use the same signaling scheme (10 kbps FSK scheme) as those in AMPS, but IS-54 voice channels use $\pi/4$ DQPSK ($\pi/4$ shifted, differential encoded quadrature phase shift keying) modulation with a raw bit rate of 48.6 kbps. The 48.6 kbps stream is divided into 6 TDMA channels of 8.1 kbps each. One IS-54 frame duration is 40 ms, and each is composed of six 6.67 ms time slots including 324 bits each. A new North-American TDMA/FDMA interim standard, IS-136 [8] provides 48.6 kbps control channel signaling on all control channels. This is the main difference between IS-54 and IS-136. IS-136 is an improved version of IS-54.

1.2.3.2 Global system for mobile communications (GSM)

GSM [3], [9] is a second generation cellular system standard that also uses TDMA/FDMA channel structure. GSM is the first cellular system to specify digital modulation and network level architectures and services, and it is the most popular cellular system in the world. The bands allocated to GSM are 890-915 MHz for the forward link and 935-960 MHz for the reverse link. The carrier spacing in GSM is 200 kHz. GMSK (Gaussian-filtered minimum shift keying) with an aggregate bit rate of 270.833 kbps is used within each 200 kHz band. Eight subscribers are time division multiplexed within the same 200 kHz band. Thus the effective channel transmission rate per user is 33.854 kbps. It should be mentioned that the GSM standard also provides for the use of frequency hopping as a means of reducing co-channel interference. Note that equalization is necessary for both IS-54 and GSM systems to compensate or reduce the signal distortion. The type of equalizer is not specified and left up to the manufacturer.

1.2.3.3 CDMA digital cellular standard (IS-95)

IS-95 DS-SS system [10] uses direct sequence spread spectrum signals with 1.25 MHz spreading bandwidth. Modulation and channel coding schemes are different for the forward link and reverse link. The basic user data rate is 9.6 kbps. The data stream is segmented into 20 ms blocks and then convolutionally encoded to provide error-correction capability. The convolutional encoder has a constraint length $k=9$ and code rate $1/2$ for the downlink and $1/3$ for the uplink. On the downlink, encoded information is further block interleaved, first spread using one of 64 Walsh sequences, and finally spread by a pair of base-specific PN sequence of length $2^{15} - 1$. Pilot-aided coherent detection is

supported by the IS-95 on the forward link. On the uplink, the data is encoded using 1/3 convolutional code, interleaved, mapped onto one of 64 Walsh sequences, and spread using a user-specific PN sequence of length $2^{42} - 1$. Note that Walsh sequences are used for different purposes on the uplink and downlink. On the downlink, Walsh codes are used for spreading to define a particular user channel, and on the uplink, they are used as 64-ary symbols in an orthogonal modulation scheme. The PN sequences in IS-95 are generated by a maximum-length shift-register.

Distinct features of the IS-95 CDMA system include:

- All users use the same carrier frequency, and hence, frequency planning, otherwise essential in cellular systems, is not necessary. For example in Fig. 1.4, all cells can share the same frequency band.
- Assignment of a single carrier frequency to all users increases the effective frequency reuse in a cellular system, and hence increases the overall capacity of the system.
- The chip duration is usually much smaller than the channel delay spread. Thus multipath components of the received signal delayed in time by more than one chip duration can be separately despread and re-combined in a RAKE combining receiver [11]. The re-combining of multipath signal components by the RAKE at the same time reduces detrimental signal dispersion and mitigates fading through the multipath diversity combining.
- Soft handoff of mobile users from one cell to another allows both the original cell and a new cell to serve the call during the handoff transition. This greatly minimizes the probability of dropping a call.

These advantages have already been validated with successful deployment of IS-95 cellular systems. CDMA has also gained support as a multiple access technique for implementation in third generation (3G) wireless communication systems.

1.2.4 Third generation cellular standard

In the first and second generation mobile cellular systems, the main design effort was primarily focused on increasing system capacity for voice services. However, the present research effort has turned toward achieving high-bit rate reliable wireless transmission of multiple services such as voice, data, facsimile etc. in the third generation of mobile cellular telecommunications [12]- [19]. As far as the multiple access technique is concerned, even though there is some interest and work on the development of 3G technologies based on TDMA [15], considerable research effort worldwide has been focused on CDMA and different CDMA systems are considered for 3G in North America, Asia and Europe.

Two major versions of CDMA 3G systems have been proposed so far. One is based on the evolution of IS-95, the so-called wideband IS-95 (or CDMA2000) [20], [25], and the other is the European/Japanese wideband direct-sequence CDMA [21], [26], [27], [28].

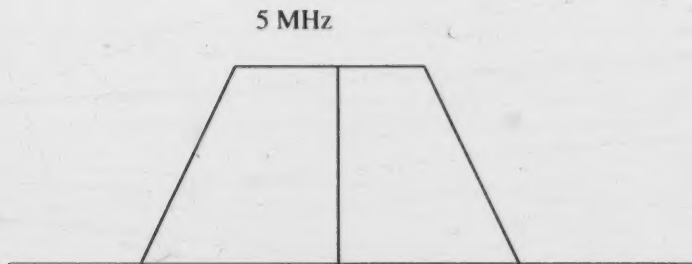
The possible frequency allocations for the wideband IS-95 are 5, 10, 15 or even 20 MHz. Some of the enhancements considered for the wideband IS-95 are as follows.

- On the reverse link, high data rate transmission is achieved by using variable spreading gain and multicode. In IS-95, the system processing gain is fixed at 128.

- Pilot-aided coherent detection technique will be used on the reverse link. The IS-95 system uses 64-ary orthogonal modulation on the reverse link.
- Two different options are proposed for the downlink (see Fig. 1.5):
 - multicarrier option with three narrowband DS-SS signals on each carrier (3X1.228 Mc/s).
 - single carrier option with wideband DS spreading (3.6864 Mc/s).
- Fast closed loop power control will be used on both reverse and forward links.



Multicarrier option



Direct spread option

Fig. 1.5. Illustration of multicarrier and direct spread transmissions

The European/Japanese wideband CDMA was originally based on the proposals in [17] and [18]. In general, the European/Japanese wideband CDMA has tried not to keep backward compatibility with any existing cellular standards. Compared with IS-95, the frame length is much shorter (10 ms) and the closed loop power control is faster (1.6 kHz). Table 1.1 and Table 1.2 summarize the key technical parameters of the European wideband CDMA and wideband IS-95 [15], [16], [20], [21].

Table 1.1. Wideband IS-95 parameters

Duplexing method	FDD
Channel bandwidths	1.25, 5, 10, 20 MHz
Forward link RF channel structure	Direct spread or multicarrier
Chip rate	1.2288/3.6864/7.3728/14.7456 Mcps
Frame length (traffic channels)	20 ms
Modulation & spreading	QPSK with quadrature spreading
Coherent detection	Continuous pilot channel
Physical control channel	Code multiplexed
Power control	Open loop and fast closed loop (800 Hz)
Multiple data rates	Variable spreading factor and multicode
Spreading	Walsh + Long codes
Coding	Convolutional codes, turbo-codes, concatenated codes
BS synchronization	Synchronous

Table 1.2. European/Japanese wideband CDMA parameters

Duplexing scheme	FDD or TDD
Carrier spacing	Flexible in the range 4.4-5.2 MHz
Chip rate	4.096/8.192/16.384 Mcps
Frame length	10 ms
Modulation & spreading	European: Balanced QPSK (downlink); Dual channel QPSK (uplink) Japanese: QPSK(downlink);BPSK (uplink)
Physical control channel	Time multiplexed (downlink); I&Q multiplexed (uplink)
Coherent detection	Time multiplexed pilot
Power control	Open and closed loop (1.6 kHz)
Multirate scheme	Variable spreading factor and multicode
Channel coding	Convolutional coding (rate $\frac{1}{2}$ - $\frac{1}{3}$), optional outer Reed-Solomon coding (rate $\frac{4}{5}$), turbo-codes
Inter-base station synchronization	FDD mode: no accurate synchronization needed; TDD mode: synchronization needed

1.3 Thesis objectives

The focus of this thesis is on the IS-95 based data transmission system. From the previous section, it can be seen that, although second generation digital cellular systems offer large

improvements in capacity and system performance over analog systems, they are designed mainly for voice service, with some provisions for low data rate services. The eventual goal of wireless communication systems is to provide subscribers with multiple services such as voice, data, facsimile, electronic mail, etc. The upcoming 3G systems are supposed to provide a single set of standards that can meet aforementioned wireless applications and universal global access, and obviously integration of voice with data is a must for 3G systems which will evolve from mature second generation systems. However, development of integrated voice/data systems creates new design challenges. Since the traffic characteristics and the quality of service requirements for data services are quite different from those of voice traffic, complete re-design of essential elements of the system is necessary to achieve efficient integration. For high data rate transmission in CDMA systems, new challenges include not only the design of new modulation, channel coding, channel estimation, and power control schemes, but also the development of new radio resource management algorithms. In this thesis, new access schemes, physical layer plus radio link protocols necessary for data transmission, call admission control policies and optimal power allocation are discussed. Specifically, the following issues are addressed:

- CDMA is an interference limited system. Packet-type services are designed for on-off sources, which generate no information during often prolonged off intervals. During off period, an approach using low bit rate maintenance signals to maintain synchronization of spreading sequences is proposed [19]. It unfortunately suffers from increased multi-user interference caused by continuously transmitted maintenance signals. Discontinuous transmission (no maintenance signals are

transmitted during the off period) is preferred in order to reduce multi-user interference. New access schemes and suitable PN sequence synchronizers are proposed and analyzed in Chapter 2.

- TCP/IP [22] has been widely used in computer networks. However TCP/IP is mainly designed for the wireline environment where the link layer frame error rate (FER) is very small (typically less than 10^{-9}). If TCP/IP is directly applied to the wireless environment where the typical value of FER is around 1% [23], it may suffer severe system performance degradation. Hence an additional link level protocol, radio link protocol (RLP) is introduced. Analysis and simulation of a complete CDMA data transmission including physical layer and retransmission protocols are presented in Chapter 3. In addition, the optimum signal-to-interference ratio (SIR) assignment is also investigated in this chapter.
- Chapter 4 presents capacity analysis of a CDMA communication system when both continuous and discontinuous transmission schemes are employed. The data messages considered include long ones for which closed loop power control can be implemented, and short ones for which only open loop power control is feasible. Call admission control (CAC) policies for a system supporting multi-rate data services are also addressed in this chapter. When an ARQ protocol is employed for data transmission, its impact on the CAC is investigated.
- Optimal forward link power allocation is considered in Chapter 5. For voice only CDMA systems, retransmission is not possible because of the critical signal delay limitation. Different users of a given cell are allocated different power levels according to their relative locations, and also their quality of service (QoS). There

does not exist a problem of optimal power allocation for voice communications. For data transmission however, due to the requirement of retransmission, there is an optimal power allocated to a particular user, which minimizes the effective total transmitted power from a base station.

Conclusions and suggestions for future work are summarized in Chapter 6.

1.4 Thesis contributions

The following original contributions are included in this thesis.

- A new access scheme, spread TDMA technique with minimum slot length is proposed. A distinct feature of this technique is that all transmitters are assigned the same spreading sequence. By doing that, only a single matched filter or correlator is required at the base station receiver to serve all packet data users.
- Performance of the digital matched filter chip synchronizer for rapid synchronization of packet data transmission has been evaluated.
- Analysis of the average access delay and throughput of the proposed access scheme in the multipath fading channel for discontinuous packet data transmission has been conducted. The proposed access system includes access messages, access schemes and fast chip synchronizers.
- Analysis and simulation of transceiver processing and retransmission protocols are presented. A complete simulation system including BCH encoding, Viterbi decoding algorithms, RAKE receiver, multipath fading channels, CDMA fast power control algorithms, etc. has been written in order to evaluate the CDMA data transmission system's performance.

- Optimum signal-to-interference ratio (SIR) assignment is investigated, and its value is obtained for a particular CDMA data transmission system.
- CAC algorithms considering the impact of ARQ protocols are proposed and capacity analysis of a multi-service CDMA system has been conducted.
- Optimal forward link power allocation has been characterized.

References

- [1] W. R. Young, "Advanced mobile phone service: Introduction, Background, and Objectives," *Bell Systems Technical Journal*, Vol. 58, January 1979, pp. 1-14.
- [2] V. H. MacDonald, "The cellular concept", *Bell Systems Technical Journal*, Vol. 58, January 1979, pp. 15-43.
- [3] K. Raith, and J. Uddenfeldt, "Capacity of digital cellular TDMA systems," *IEEE Transactions on Vehicular Technology*, Vol. 40, No. 2, pp. 323-331, May 1991.
- [4] K. S. Gilhousen, etc., "On the capacity of a cellular CDMA system," *IEEE Trans. on Vehicular Technology*, Vol. 40, No. 2, pp. 303-312, May 1991.
- [5] R.L. Peterson, R.E. Ziemer and D.E. Borth, *Introduction to Spread Spectrum Communications*, Prentice-Hall, New Jersey, 1995.
- [6] A.J. Viterbi, *CDMA: Principles of Spread Spectrum Communication*. Addison-Wesley, 1995.
- [7] EIA/TIA Interim Standard IS-54, "Cellular system dual mode mobile station - base station compatibility specifications," *Electronics Industry Association*, May 1990.
- [8] J. Padyett, E. Gunther, and T. Hattäri, "Overview of wireless personal communications," *IEEE Communications Magazine*, January 1995.

- [9] M. Mouly and M. B. Pautet, "The GSM system for mobile communications", published by the authors, 1992.
- [10] TIA/EIA IS-95, "Mobile Station-Base Station Compatibility Standard for Dual-Mode Wideband Spread-Spectrum Cellular Systems", *Telecommunications Industry Association*, July 1993.
- [11] J. Proakis, *Digital communications*, 3rd edition, McGraw-Hill, Inc., 1995.
- [12] A. Fukasawa, T. Sato, Y. Takizawa, T. Kato, M. Kawabe and R. E. Fisher, "Wideband CDMA system for personal radio communications", *IEEE Communications Magazine*, October 1996, pp. 116-123.
- [13] K. Letaief, J. C.-I. Chuang and R. D. Murch, "A high-speed transmission method for wireless personal communications", *Wireless Personal Communications*, 3: pp 299-317, 1996.
- [14] N. Morinaga, M. Nakagawa and R. Kohno, "New concepts and technologies for achieving highly reliable and high-capacity multimedia wireless communications systems", *IEEE Communications Magazine*, January 1997, pp. 34-41.
- [15] T. Ojanperä, "Development of 3G radio technologies", *First Annual UCSD Conference on Wireless Communications*, San Diego, California, March 1998, pp. 127-132.
- [16] E. G. Tiedemann Jr., Yu-Cheun Jou, and J. P. Odenwalder, "The evolution of IS-95 to a third generation system and to the IMT-2000 era", *ACTS Mobile Communications Summit '97*, Aalborg, Denmark, October 1997, pp. 924-929.

- [17] F. Ovesjö, E. Dahlman, T. Ojanperä, A. Toskala and A. Klein, "FRAMES multiple access mode 2 – wideband CDMA", *Proceedings of IEEE PIMRC97*, Helsinki, Finland, September 1997, pp. 42-46.
- [18] Y. Ishida, "Recent study on candidate radio transmission technology for IMT-2000", *First Annual CDMA European Congress*, London, UK, October 1997.
- [19] A. Baier, "Open multi-rate radio interface architecture based on CDMA," *Proc. IEEE ICUPC'93*, Ottawa, Canada, Oct. 1993, pp. 985-989.
- [20] TR45.5.0.18, "The CDMA2000 ITU-R RTT Candidate Submission", June 1998.
- [21] ETSI SMG2, "The ETSI UMTS Terrestrial Radio Access (UTRS) ITU-R RTT Candidate Submission", June 1998.
- [22] A. Tanenbaum, *Computer Networks*, 3rd edition, Prentice-Hall, New Jersey, 1996.
- [23] A.M. Viterbi and A.J. Viterbi, "Erlang capacity of a power controlled CDMA system", *IEEE Journal on Selected Areas in Communications*, Vol. 11, No. 6, August 1993, pp. 892-900.
- [24] T.S. Rappaport, *Wireless Communications: Principles and Practice*, Prentice-Hall, New Jersey, 1996.
- [25] D.N. Knisely, S. Kumar, S. Laha and S. Nandy, "Evolution of wireless data services: IS-95 to CDMA2000", *IEEE Communications Magazine*, vol. 36, no. 10, October 1998, pp. 140-149.
- [26] F. Adachi, M. Sawahashi and H. Suda, "Wideband DS-CDMA for next generation mobile communications systems", *IEEE Communications Magazine*, vol. 36, no. 9, September 1998, pp. 56-69.

[27] E. Dahlman, B. Gudmundson, M. Nilson and J. Sköld, "UMTS/IMT-2000 based on wideband CDMA", *IEEE Communications Magazine*, vol. 36, no. 9, September 1998, pp. 70-80.

[28] T. Ojanperä and R. Prasad, "An overview of air interface multiple access for IMT-2000/UMTS", *IEEE Communications Magazine*, vol. 36, no. 9, September 1998, pp. 82-95.

Chapter 2

DESIGN AND PERFORMANCE EVALUATION OF ACCESS SCHEMES

2.1 Introduction

The 3G CDMA system is supposed to support simultaneous voice/packet data/circuit data operations [1] with different QoS requirements. Since the traffic characteristics and the quality of service requirements for packet data services are quite different from those of voice traffic, system re-design in certain essential aspects is necessary for an integrated voice-data system without impacting voice quality or sacrificing data performance.

Packet-type services are designed for on-off sources, which generate no information during often prolonged, but unpredictable, off intervals. For many data applications the intermittent nature of the traffic is orders of magnitude greater than that of voice signals (Fig. 2.1). The system capacity can be greatly increased if no signal is transmitted during the off period, since the CDMA system is interference limited.

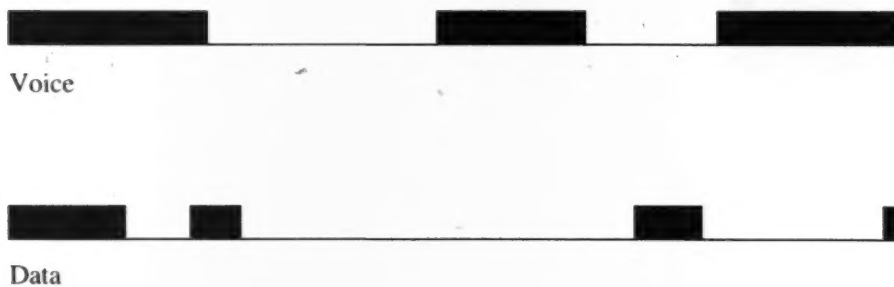


Fig. 2.1. Traffic characteristics of voice and data.

The efficient provision of packet-type services is particularly difficult on reverse links (from portable to base station) of CDMA systems, because of the need to rapidly regain synchronization of spreading sequences when the source resumes transmission after a period of silence. One possible solution to the link maintenance problem when bursty transmission is required on reverse links of CDMA systems is to assign a separate low bit rate physical control channel (created using code division multiplex) to each portable in a given cell. This approach, proposed by some investigators [2], can be called a continuous transmission medium access control scheme (CTX-MAC), and it unfortunately suffers from increased multi-user interference caused by continuously transmitted maintenance signals. Increased interference naturally leads to reduced traffic capacity. Also its implementation is characterized by significantly increased complexity of base station receivers, where each physical control channel (serving one portable) would require a separate despreading correlator. If the duty cycle (ratio of the on-period to off-period of the mobile's transmitter) as shown in Fig. 2.1 is small, then considerable savings in both base station hardware and system capacity may be achieved if transmission is discontinued during the off-periods and the hardware is shared among different users. This is the discontinuous transmission medium access control (DTX-MAC) scheme. Obviously, with the DTX-MAC scheme, when a particular user has a data packet to send, the base station needs to be notified by the mobile of its intention to transmit, and synchronization needs to be quickly resumed. To achieve that an access request message (ARM) is transmitted from the mobile to base station to acquire synchronism, and to inform the base station of the mobile's identity. An access reservation channel (ARC) is

allocated for such a purpose. All mobiles in a cell (or sector) use the same PN code to send their ARMs on the ARC; this avoids the need to have a separate receiver for each mobile, which is in the off mode. If the number of sources is large, then more than one ARC may be used.

A MAC protocol is required on the ARC for the mobiles to access the base station with individual ARMs. One possible approach is to allow the mobile units to send ARMs in an asynchronous fashion - ADTX-MAC protocol. Another is a synchronous approach - SDTX-MAC protocol, in which a slotted structure is used for the access control of ARMs from all users. An improvement on the latter scheme may allow for an overlap in time of packets transmitted by different users.

Since the user sending a transmission request has already been registered and has acquired synchronization before, the timing ambiguity is mainly caused by the propagation delay uncertainty due to the movement of the mobile during the silent period. Hence, it is relatively easier to synchronize again than to acquire initial synchronization during the registration phase involving the access channel. On the other hand, the synchronization needs to be regained very quickly so that the arriving data burst can be transmitted without undue delay. Due to these reasons, the specific structure of the IS-95 access channel [3] which is used by the mobile station to initiate communication with the base station and to respond to paging channel messages, is no longer suitable, since its access probe occupies several 20 ms frames in time. Instead, much shorter access request

messages are used within the access reservation channel structure significantly different from the IS-95 access channel. Such an approach dramatically reduces access delay and increases throughput.

The access system proposed in this chapter is composed of three parts. The first part is the access protocol, the second is the access request message (ARM), and the last is the synchronizing circuit.

2.2 Proposed Access Protocols.

2.2.1 SDTX-MAC protocol¹

A time-slotted frame structure is used to accommodate access requests from different users on the access reservation channel (ARC). Specific time slots in the frame equal in length to the duration of the access request message (if necessary, a small guard time may be added) are assigned by the base station to mobile users after registration (one slot per user). Collision is thus avoided and there is no need of inserting an identifier appendix after the synchronizing part of the ARM. A method of step-increase of power is used for ARMs. After each ARM, the mobile monitors the paging channel for an acknowledgment (ACK) from the base station. If the elapsed time before the ACK is received is longer than the prescribed maximum, the corresponding ARM is regarded as a failure, and next one is sent at a power level increased by a fixed step. For each user, a fixed number of ARMs of increasing powers form an ARM sequence. If the synchronization is still not acquired after the sequence of ARMs is sent, a new sequence is started, with the

¹ Patent pending.

transmitting power starting from the lowest level. Identical ARM sequences are repeated up to a maximum allowed number of repetitions, which is assumed large enough to ensure acquisition. The synchronism acquired through the ARC may be coarse, but it is sufficiently accurate to receive data packets from the user in question on a data channel. Each data packet sent on that channel is preceded by a short synchronizing preamble to refine synchronism.

The number of slots in one slotted frame of the access reservation channel (ARC) is designed to accommodate the expected number of registered users in a cell or sector. Once a user is registered, a slot is assigned to him on the ARC, and the base station subsequently uses that assignment to identify a user sending an access request. An example is shown in Fig. 2.2 wherein there are four slots within one frame, T_p , T_F are the slot and frame lengths, respectively, and there are three ARMs within one access message sequence.



Fig. 2.2. Structure and parameters of the SDTX-MAC protocol.

2.2.2 ADTX-MAC protocol.

The asynchronous medium access control protocol is in general terms similar to that used in the IS-95 access channel. It is a spread slotted ALOHA [4]-[6] with p -persistence [7] (plus sequence back-off) between ARM sequences. As shown in Fig. 2.3 a fixed number of ARMs forms an ARM sequence (three ARMs within an ARM sequence in this figure), with the transmission power increasing consecutively by a fixed increment. There is a random back-off T_R between ARM sequences, if the ACK is not received before time-out, T_A . If the second ARM sequence is required, there is a random back-off T_S before a p -persistence test, for the next ARM sequence to start. To reduce the probability of collision, the exact transmission time of each user is pseudo-randomly delayed from the slot start time, and the delay is known to the base station.

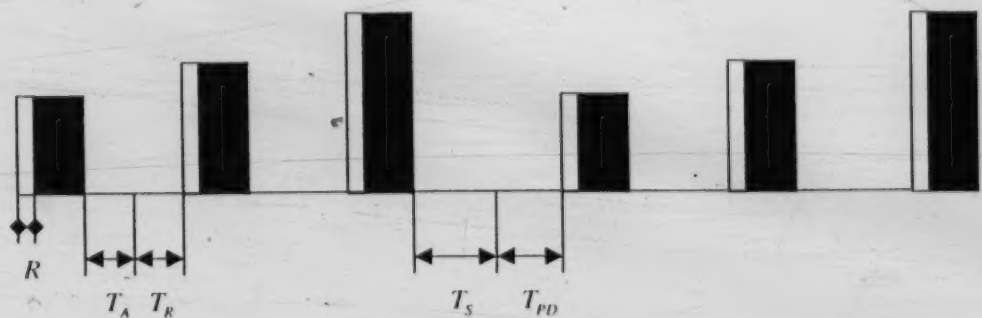


Fig. 2.3. Structure and parameters of the ADTX-MAC protocol.

Since all the users use the same PN code to spread their ARMs, each ARM consists of one synchronization preamble and an appendix for user identification. This appendix is designed to have the same length as the preamble, and is block-coded to provide the

necessary error protection. It is noted that the ARM for ADTX-MAC is of twice the length of an ARM in SDTX-MAC protocol.

2.2.3 HDTX-MAC protocol.

The hybrid DTX-MAC protocol is similar to the SDTX-MAC. The main difference between the HDTX-MAC (illustrated in Fig. 2.4) and the SDTX-MAC is that more than one user is allowed to transmit packets within a longer time slot, and each longer slot is divided into short minislots, each of which is dedicated to one of the mobile terminals. Packets are transmitted at the beginning of each minislot. As it will be demonstrated in a later section, T_{LP} (duration of the extended slot) is only moderately longer than T_P (duration of a single-user slot), since the minislot duration T_m is very short compared with T_P .

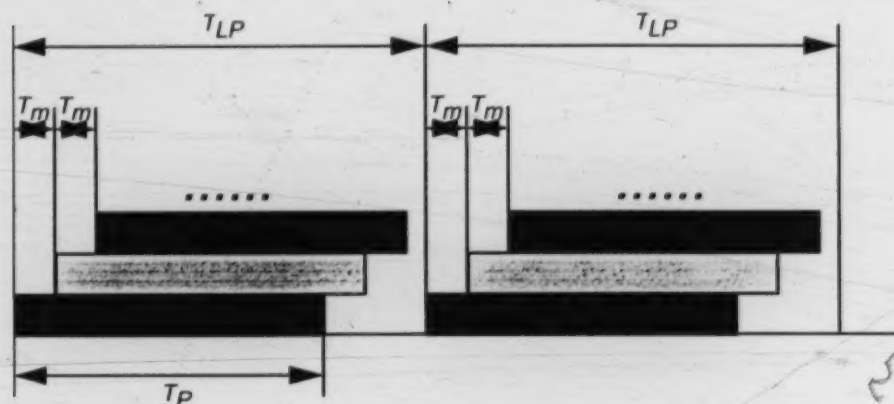


Fig. 2.4. Structure and parameters of the HDTX-MAC protocol.

2.3 Structure of Access Request Messages

In the SDTX scheme the ARM contains a number of repetitions of a short PN code. The

ARM in the HDTX-MAC scheme is exactly the same as that of the SDTX scheme. In the ADTX-MAC scheme an ARM contains a synchronizing preamble which is exactly the same as the ARM in the SDTX-MAC, and an appendix for identification. The ID appendix contains sixteen 64-ary Walsh symbols and each Walsh symbol will be further spread by a factor of 4. Each Walsh symbol in the appendix contains six codeword elements. A shortened binary (96, 33) BCH code is used to encode the appendix. The code is shortened from the binary primitive (127, 64) BCH code with $d_{\min}$ equal to at least 21 [8]. Since a shortened cyclic code has error detection and correction capability at least as good as that of the original code, the (96, 33) BCH code can detect all error patterns containing up to 20 errors, and it uses the same encoder and decoder as the original (127, 64) BCH code.

2.4 Synchronizer Circuits

2.4.1 Background

The process of synchronizing the received PN signals is carried out in two steps. The first step is the determination of the initial code phase; this is called code acquisition. The purpose of the code acquisition is to align the two PN signals within a small (typically less than $\frac{1}{2}$ of a PN chip) relative timing offset. The second step is to achieve fine code synchronization after initial acquisition, which is called code tracking. Only code acquisition is discussed here.

The most commonly used PN sequence acquisition technique is the so-called serial-

search technique, in which the synchronization system evaluates the time cells of Fig. 2.5 one after another (serially) until the correct cell is found. For example, as shown in Fig. 2.5, if the system first evaluates cell #1, it will fail since the correct cell is #5; the serial-search system will keep trying cell #2, then #3, and finally hit the correct cell #5.

☺ denotes correct phase

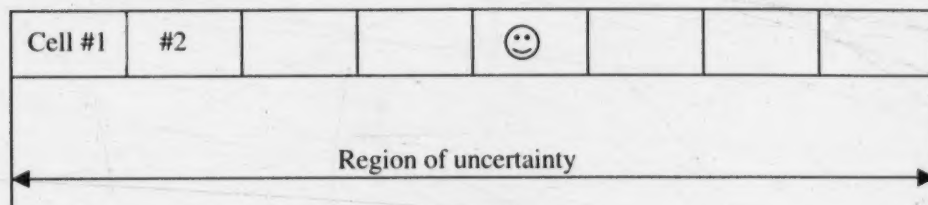


Fig. 2.5. Time or phase uncertainty region.

Since the initial code phase is commonly determined before any carrier phase measurement or tracking is attempted, the noncoherent detector must be used in the acquisition system. That is, the synchronizer might be comprised of a bandpass filter centered at the carrier frequency, followed by a square-law envelope detector, an integrate-and-dump circuit, and a threshold device [9], [10]. Depending on whether the detector's decision is made on the basis of a single fixed time observation of the received signal or many such observations, commonly used serial search techniques can be further divided into single dwell and multiple dwell types [11]-[16]. A simple block diagram of the single dwell acquisition system is illustrated in Fig. 2.6 [9], where $r(t)$ is the received signal, f_0 is the carrier frequency, $H'(f)$ is the transfer function of a filter matched to

the PN chip waveform, and τ_d is the fixed integration time or dwell time.

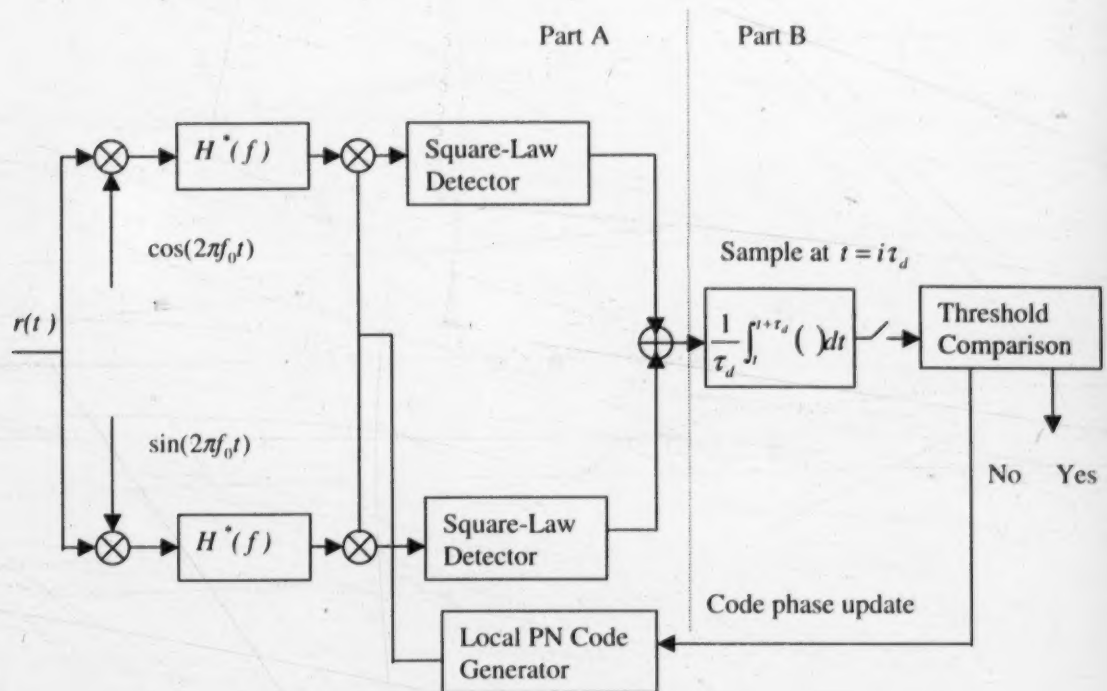


Fig. 2.6. Block diagram of the single dwell time PN acquisition system.

The acquisition procedure for the single dwell scheme in Fig. 2.6 can be briefly described as follows: the received signal is first down-converted to in-phase and quadrature components, then passed through the chip waveform matched filter, and actively correlated with the local replica of the PN code. The output is then square-law envelope detected and integrated over a fixed dwell duration, τ_d , in an integrate-and-dump circuit, and then compared to a preset threshold. If the output is larger than the threshold, a "hit" or correct cell is declared found, and if not, the local PN code generator will update its

code phase (go to another cell), and the search will continue until the correct cell is found.

In a multiple dwell serial search code acquisition system, the examination interval consists of a series of fixed short dwell periods (each longer than its predecessor) with a decision being made after each. This technique is illustrated in Fig. 2.7. Note that the part A in this figure is exactly the same as Fig. 2.6.

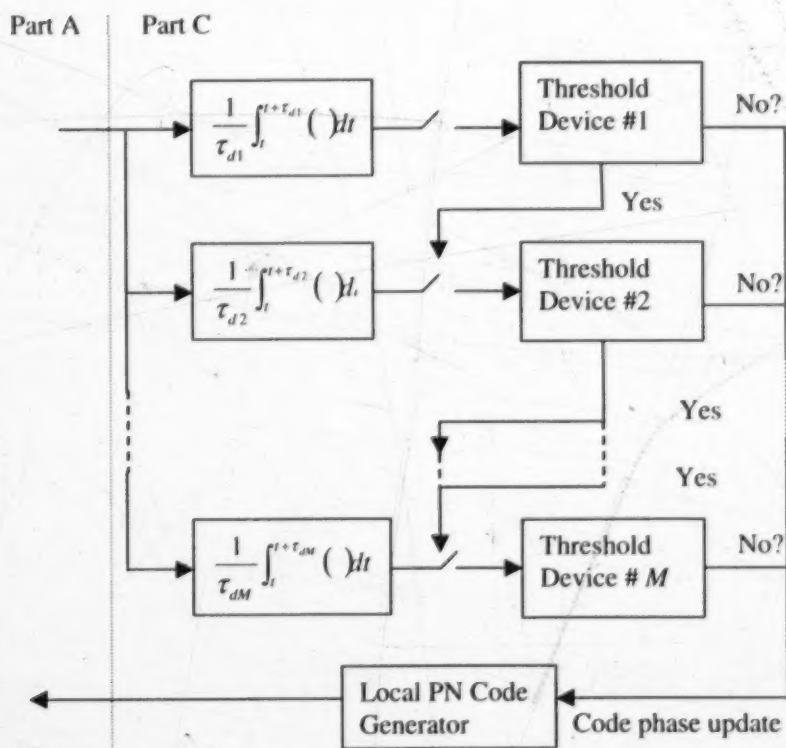


Fig. 2.7. Block diagram of the multiple dwell time PN acquisition system.

In Fig. 2.7, the multiple dwell periods are ordered such that:

$$\tau_{d1} \leq \tau_{d2} \leq \dots \leq \tau_{dM} \quad (2.1)$$

The principle of operation for the front end of the system ("part A" in the Figure) is the same as that in Fig. 2.6. Differently from the single dwell search scheme, a multiple dwell search scheme makes a decision by examining all M threshold devices in Fig. 2.7. Specifically, a multiple dwell search is carried out according to the following algorithm:

1. The output from the front end is integrated for τ_{d1} . At the end of the integration, if the result is larger than the threshold, a "hit" is declared and a second integration begins. If a second "hit" is declared, then the third integration begins. Finally, if all of the M threshold devices indicate the correct cell is hit, the decision is made and the search will stop.
2. If any one threshold device shows no threshold crossing, a "miss" is declared, and local code generator will update its code phase, upon which the next cell is examined. In other words, as soon as one device indicates the codes are misaligned, the remaining devices will stop making decisions.

In all of the code acquisition schemes discussed above, the received signals were correlated with a locally generated replica of the spreading waveform to determine whether the local phase is correct. The duration of this correlation (τ_d in Fig. 2.6) may be anywhere from 10 to several thousand times of the chip duration [15]. Denote $\tau_d = KT_c$, where T_c is the duration of a spreading chip. Since the local PN code reference phase can be updated only at τ_d intervals, the search rate is [9]:

$$\mathfrak{R} = 1/\tau_d = 1/KT_c. \quad (2.2)$$

The search rate can be significantly increased if the time to evaluate each phase cell is reduced. A matched filter can be used to replace the correlation-and-integration operation as shown in Fig. 2.8 [9], [15], where a baseband filter is matched to a segment of the PN spreading waveform.

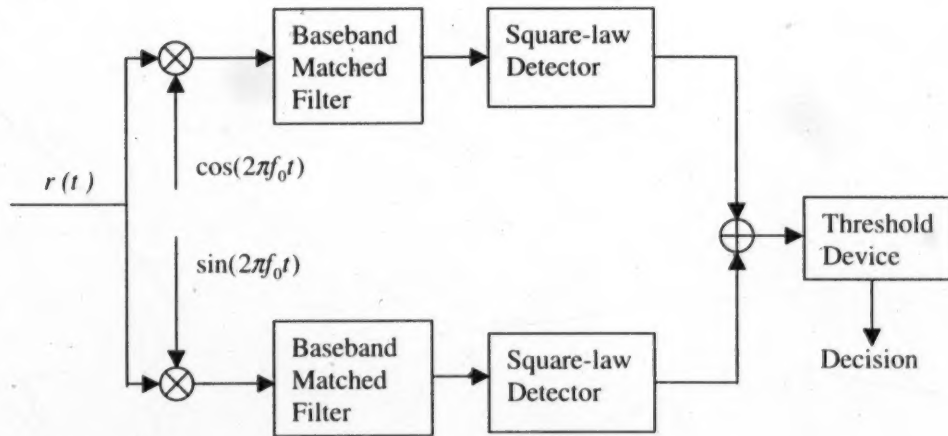


Fig. 2.8. A matched filter acquisition synchronizer.

Conceptually, the filter matched to a segment of the PN spreading waveform can be realized as a cascade of a filter matched to the chip waveform followed by a digital filter matched to the spreading PN code. Suppose now that $s(t)$ corresponds to an N -chip segment of the PN waveform. If N is the period of the PN code, we say a full period PN code correlation is used; otherwise, a partial PN code correlation is utilized. We have the following expression for $s(t)$:

$$s(t) = \sum_{i=1}^N a_i h(t - iT_c) \quad (2.3)$$

where $\{a_i\} = \{a_1, a_2, \dots, a_N\}$ is a PN code vector, $a_i \in \{-1, 1\}$, and $h(t)$ is the basic chip waveform which in the simplest case is given by:

$$h(t) = \begin{cases} \frac{1}{\sqrt{T_c}}; & 0 \leq t \leq T_c \\ 0; & \text{otherwise} \end{cases} \quad (2.4)$$

Equation (2.3) can also be written as:

$$s(t) = h(t) * \sum_{i=1}^N a_i \delta(t - iT_c) \quad (2.5)$$

where $*$ denotes the convolution operation, and $\delta(t)$ is the Dirac delta function. The Fourier transform of $s(t)$ is:

$$S(f) = H(f) \sum_{i=1}^N a_i e^{-j2\pi f iT_c} \quad (2.6)$$

where $j = \sqrt{-1}$. The transfer function of the filter matched to $s(t)$ is:

$$S_M(f) = S^*(f) e^{-j2\pi f NT_c} = H^*(f) \sum_{i=1}^N a_i e^{j2\pi f iT_c} e^{-j2\pi f NT_c} = H^*(f) \sum_{i=1}^N a_{N-i+1} e^{-j2\pi f (i-1)T_c} \quad (2.7)$$

where superscript $*$ denotes complex conjugation. Clearly (2.7) can be implemented in the form of a filter matched to a single PN chip waveform followed by a discrete-time filter matched to the PN code, which is illustrated in Fig. 2.9. The sampling period T_s in the discrete-time implementation may be shorter than the chip interval: $T_s = T_c / k$, where $k = 1, 2, \dots$. The discrete-time matched filter in Fig. 2.9 then takes a form depicted in Fig. 2.10, where b_l is the l -th sub-chip polarity related to a_i in Fig. 2.9 by:

$$b_{(i-1)k+l} = a_i \text{ for } l = 1, 2, \dots, k; i = 1, 2, \dots, N$$

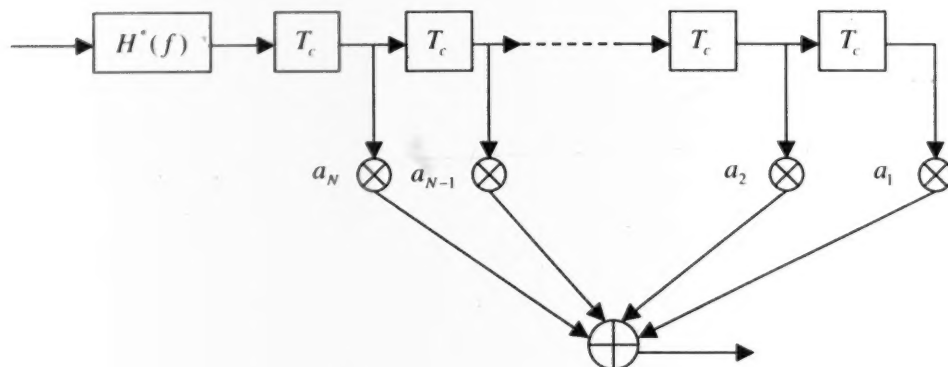


Fig. 2.9. Discrete-time implementation of a filter matched to an N -chip spreading waveform.

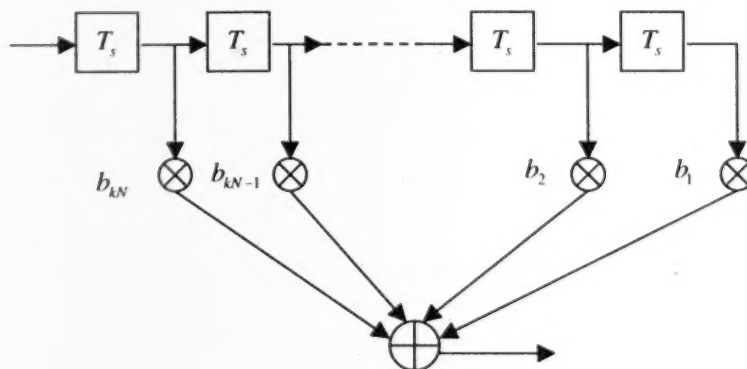


Fig. 2.10. Oversampled discrete-time matched filter.

The code phase in Fig. 2.8 is tested after each interval T_s . In other words, only one new fraction of $1/k$ chip of received signal is used for each correlation test to determine if the received code phase is correct; note that this new signal is combined with the prior $kN - 1$

received signal samples stored in the shift register. Thus, the search rate for this acquisition method is clearly [9]:

$$\mathfrak{R} = 1/T_s = k/T_c \quad (2.8)$$

Comparing (2.8) with (2.2) we see that the acquisition technique using a matched filter is a factor of $(k/T_c)/(1/KT_c) = kK$ faster than that of the serial search technique. This reduction in search time has been achieved by freezing the phase of the receiver reference spreading code and waiting for a particular phase of the received waveform to arrive. The penalty paid for the matched filter search technique is that the receiver needs to store prior received information before testing each new received signal. The serial search scheme may be called a search technique without memory, and the matched filter technique is a search technique with memory.

Due to its short search time, a digital differential matched filter synchronizer for rapid synchronization of packet data transmissions is considered here.

2.4.2 Modified synchronizer using matched filter techniques

Fundamental digital spread spectrum modulators for both BPSK and QPSK spreading [17] are presented in Fig. 2.11 and Fig. 2.12, and the corresponding simplified block diagrams of the synchronizers for both BPSK spreading and QPSK spreading are shown in Fig. 2.13 and Fig. 2.14.

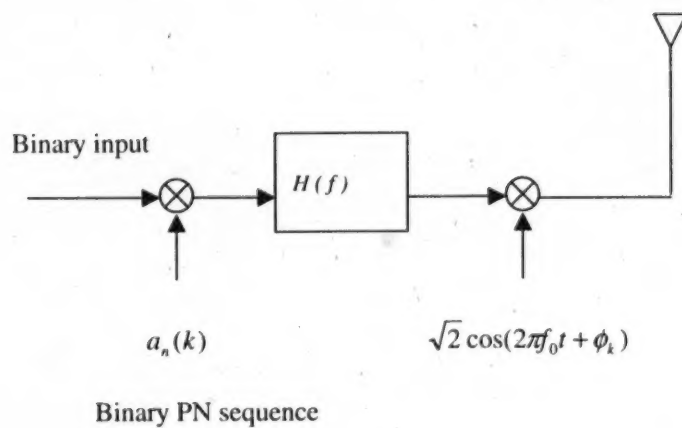


Fig. 2.11. BPSK direct sequence spread spectrum transmitter.

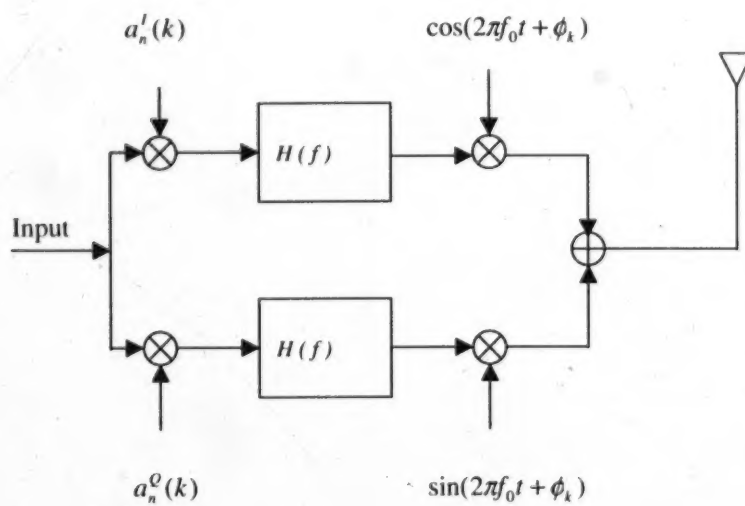


Fig. 2.12. QPSK direct sequence spread spectrum transmitter.

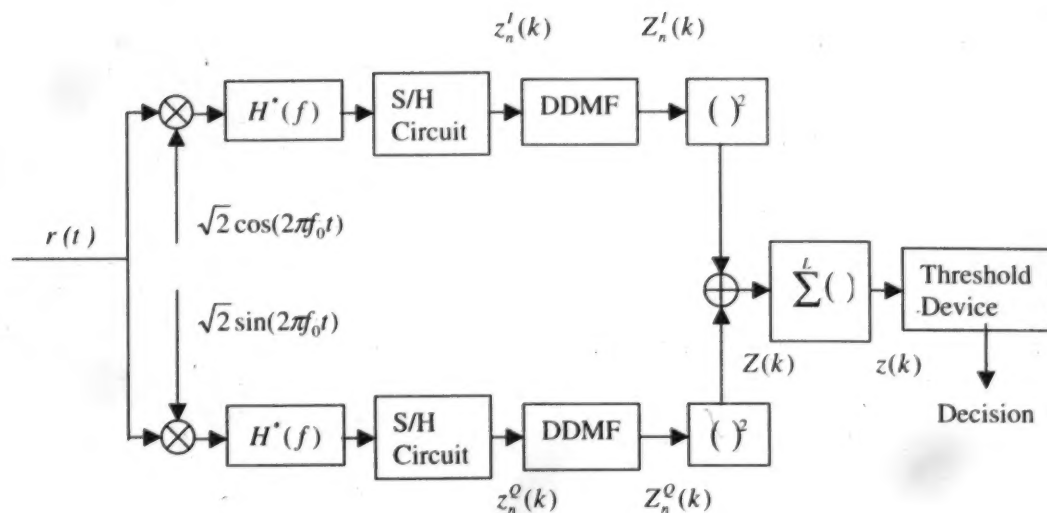


Fig. 2.13. Block diagram of the synchronizer for BPSK spreading.

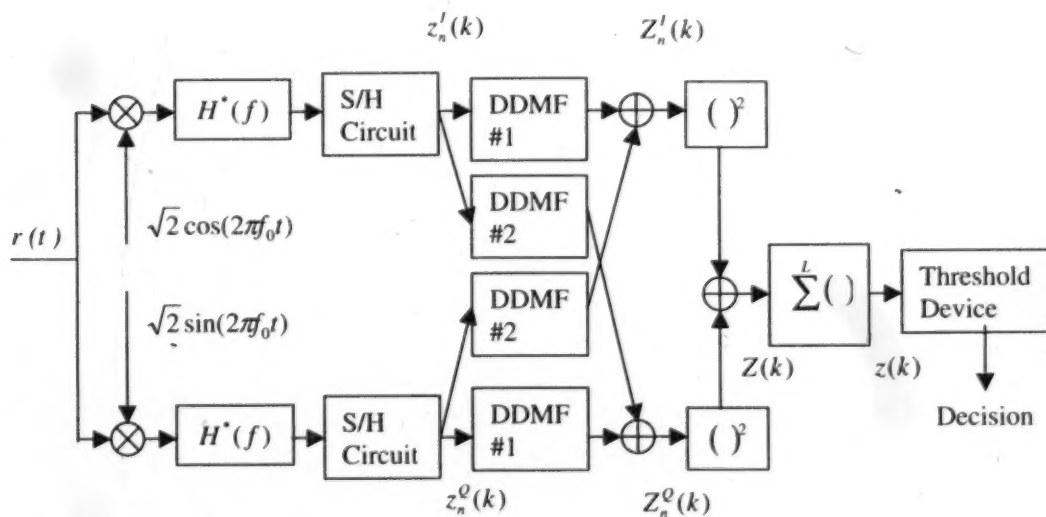


Fig. 2.14. Block diagram of the synchronizer for QPSK spreading.

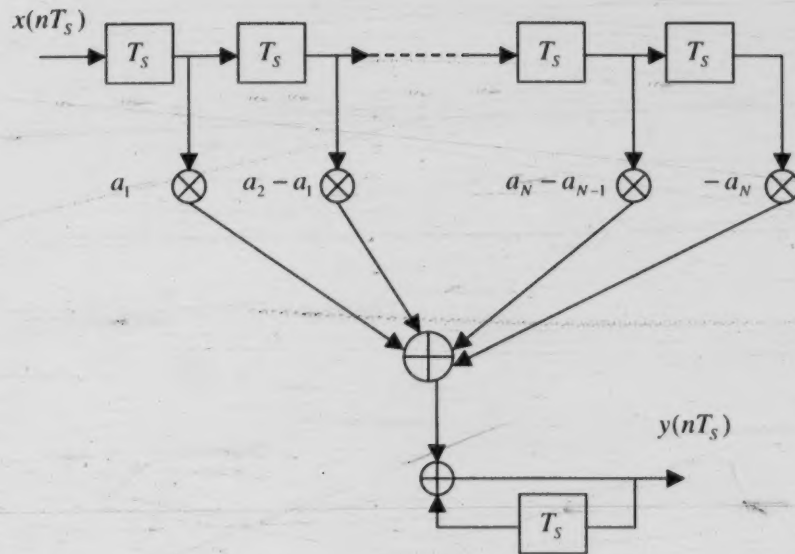


Fig. 2.15. Differential discrete-time matched filter.

In Fig. 2.13 and Fig. 2.14, the sampling & hold circuit's sampling period is T_s , and DDMF refers to the differential discrete-time matched filter (see Fig. 2.15); in Fig. 2.14, DDMF #1 is matched to the in-phase PN sequences $a_n^I(k)$, and DDMF#2 to the quadrature PN sequences $a_n^Q(k)$. Similar to the operation in Fig. 2.8, the received signal is first down converted to in-phase and quadrature components, which are then passed through the waveform matched filter, sampled at the rate of $1/T_s$, and fed to the DDMF. The element for the implementation of the sequential accumulation of outputs (measurements) is added before the threshold test device. Synchronism is declared at the sampling point where the accumulated value exceeds the threshold, which is chosen to keep the probability of false acquisition below a certain value. The structure of the

DDMF proposed in [18] is shown in Fig. 2.15, where $x(nT_s)$ is the received digital signal at time nT_s , $a_1, a_2, \dots, a_N$ are the PN sequence chips, and N is the period of the PN sequence. The DDMF is initially reset to zero. The following relations hold at instants nT_s ($n=1,2,\dots,N$):

$$y(T_s) = a_1 x(T_s)$$

$$y(2T_s) = (a_2 - a_1)x(T_s) + a_1 x(2T_s) + y(T_s) = a_2 x(T_s) + a_1 x(2T_s)$$

$$y(NT_s) = a_N x(T_s) + \dots + a_1 x(NT_s) \quad (2.9)$$

From equation (2.9), we see that the output of the DDMF is equal to that of the conventional discrete-time matched filter. The difference between a conventional discrete-time matched filter and the DDMF is that the weighting taps of the former are either 1 or -1, and those of the latter are 2, 0, or -2 except for a_1 and $-a_N$. There is no need for multiplication when the taps are zero, so the number of operations is reduced.

2.4.3 Performance analysis

Probabilities of correct and false acquisitions of the chip synchronizers for both Gaussian and Rayleigh fading channels are derived in this section. The methodology employed here is similar to that used in [17].

When the BPSK spread spectrum modulator is specified as in Fig. 2.11, the received

signal in Fig. 2.13 can be written as

$$r(t) = \sqrt{2E_c(k)} d_n(k) a_n(k) h(t - nT_c) \cos(2\pi f t + \phi_k) + n(t) \quad (2.10)$$

where the index k indicates k th user, the subscript n refers to the n -th chip, $E_c(k)$ is the chip energy, $d_n(k)$ is the data bit assuming values of $+1$ or -1 , $a_n(k)$ is the binary spreading sequence element assuming values of $+1$ or -1 , ϕ_k is a random carrier phase, $n(t)$ is the AWGN with one-sided power spectral density N_0 and zero mean, and $h(t)$ is given by (2.4). According to Parseval's theorem:

$$\int_{-\infty}^{\infty} |H(f)|^2 df = \int_0^{T_c} h^2(t) dt = 1 \quad (2.11)$$

Since only synchronizing messages are transmitted during acquisition, the data values are thus always constant ones. When $T_s = T_c$, assuming a perfect chip synchronization we have (see Fig. 2.13):

$$E[z_n^I(k)] = \sqrt{E_c(k)} R(\tau) \cos(\phi_k) \quad (2.12)$$

$$E[z_n^Q(k)] = \sqrt{E_c(k)} R(\tau) \sin(\phi_k) \quad (2.13)$$

where τ is the timing error, and $R(\tau)$ is given by

$$R(\tau) = \begin{cases} 1 - \frac{|\tau|}{T_c} & |\tau| \leq T_c \\ 0 & \text{otherwise} \end{cases} \quad (2.14)$$

$z_n^I(k)$ and $z_n^Q(k)$ also include the components due to background noise, and the other-user interference components. The variance due to the background noise is given by:

$$V_{N_0} = (N_0/2) \int |H(f)|^2 df = N_0/2 \quad (2.15)$$

The interference from a user's signal spread by a PN chip waveform with chip energy E_c

and chip duration T_c can be modeled as an equivalent white noise of two-sided density E_c/T_c [17]. In order to calculate the other user interference, from Fig. 2.11 and Fig. 2.13 we may simplify the noise model of the transmitting and receiving process as illustrated in Fig. 2.16.

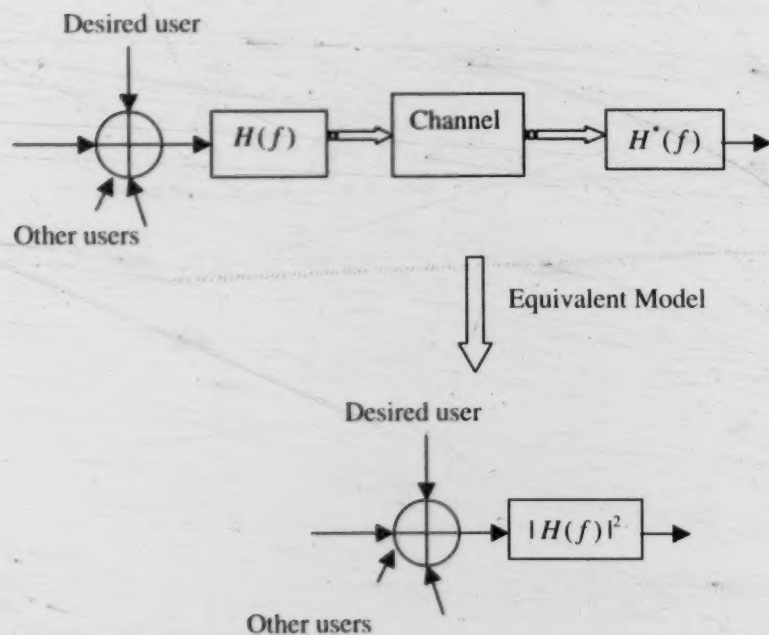


Fig. 2.16. Simplified interference model.

From Fig. 2.16, we see that by assuming undesired users' signals as white noise, the effect of the j -th user's signal on the k -th user (desired one) will be that of white noise with two-sided density $E_c(j)/T_c$ passing through the filter with transfer function $|H(f)|^2$. Taking into account the relative phases of the j -th user and k -th users, the

variance, denoted by V_{other} , due to the j -th user's signal is:

$$V_{other} = \sum_{j \neq k} [E_c(j)/T_c] \cos^2(\phi_j - \phi_k) \int |H(f)|^4 df \quad (2.16)$$

The variances of $z_n^I(k)$ and $z_n^Q(k)$ are thus given by:

$$Var[z_n^I(k)] = Var[z_n^Q(k)] = V_{N_0} + V_{Other} \quad (2.17)$$

Note that here it is assumed that $z_n^I(k)$ and $z_n^Q(k)$ are Gaussian variables.

It is reasonable to assume that the relative phases of all users are uniformly distributed between 0 and 2π . By averaging (2.16) over all phases, we have:

$$V_{other} = (1/3) \sum_{j \neq k} E_c(j) \quad (2.18)$$

where we have used the following facts:

$$E[\cos^2 \phi] = 1/2$$

$$\frac{1}{T_c} \int_{-\infty}^{\infty} |H(f)|^4 df = \frac{1}{T_c} \int_{-T_c}^{T_c} R^2(\tau) d\tau = \frac{2}{T_c} \int_0^{T_c} (1 - \frac{\tau}{T_c})^2 d\tau = 2/3$$

where $R(\tau)$ (see 2.14) and $|H(f)|^2$ are the Fourier transform pair.

Denoting $V_{N_0} + V_{other} = \frac{V_0}{2}$, we have:

$$V_o = N_0 + (2/3) \sum_{j \neq k} E_c(j) \quad (2.19)$$

Since the number of the taps in the matched filter is N , which is also the period of the synchronizing PN sequence, we can further obtain the following means and variances of the outputs from the DDMF (for simplification, the index n has been dropped in the following analysis):

$$E[Z^I(k)] = N\sqrt{E_c(k)}R(\tau)\cos(\phi_k) \quad (2.20)$$

$$E[Z^Q(k)] = N\sqrt{E_c(k)}R(\tau)\sin(\phi_k) \quad (2.21)$$

$$\text{Var}[Z^I(k)] = \text{Var}[Z^Q(k)] = NV_o / 2 \quad (2.22)$$

In case that the desired signal is absent, the mean values of both $Z^I(k)$ and $Z^Q(k)$ should be zero, but the variances should be the same as in (2.22). Since $Z(k) = [Z^I(k)]^2 + [Z^Q(k)]^2$, hence, Z is a central chi-square distribution random variable with two degrees of freedom when the signal is absent. Hence the probability density function (pdf) and the characteristic function of Z are [19]:

$$p_0(Z) = \frac{1}{NV_o} e^{-Z/NV_o} \quad (2.23a)$$

$$\psi_z^0(jv) = \frac{1}{1 - jvNV_o} \quad (2.23b)$$

When signal is present, Z is a noncentral chi-square variable with noncentrality parameter $N^2 E_c R^2(\tau)$. Note that here we have dropped the index k assuming that the chip energy for all users is E_c . The pdf and the characteristic function of Z are also given by [19]:

$$p_1(Z) = \frac{1}{NV_o} e^{-(Z + N^2 E_c R^2(\tau))/NV_o} I_0\left(\frac{2R(\tau)\sqrt{ZE_c}}{V_o}\right) \quad (2.24a)$$

$$\psi_z^1(jv) = \frac{1}{1 - jvNV_o} \exp\left(\frac{jvN^2 E_c R^2(\tau)}{1 - jvNV_o}\right) \quad (2.24b)$$

where $I_0(x)$ is the zero order modified Bessel function of the first kind.

As seen in Fig. 2.13, the final decision is based on the comparison between the sum of the L observations and a threshold. Defining the variable z as the summation, and assuming that the L measurements are independent, the characteristic function of z is the L -th power of the right side of either (2.23b) or (2.24b). The probability density functions of z are obtained by inverting the corresponding Fourier transforms, and they have the following forms [19]:

$$p_0(z) = \frac{z^{L-1} e^{-z/NV_o}}{(L-1)!(NV_o)^L} \quad (2.25)$$

$$p_1(z) = \frac{1}{NV_o} \left(\frac{z}{LN^2 E_c R^2(\tau)} \right)^{(L-1)/2} e^{-(z + LN^2 E_c R^2(\tau))/NV_o} I_{L-1} \left(\frac{2R(\tau)\sqrt{LE_c}z}{V_o} \right) \quad (2.26)$$

If the decision threshold in Fig. 2.13 is θ , we can express the false and correct detection probabilities of the synchronizer for BPSK spreading as follows [17]:

$$P_F = \int_{\theta}^{\infty} p_0(z) dz = e^{-\theta/NV_o} \sum_{k=0}^{L-1} \frac{(\theta/NV_o)^k}{k!} \quad (2.27)$$

$$P_D = \int_{\theta}^{\infty} p_1(z) dz = \int_{\theta/NV_o}^{\infty} \left(\frac{x}{L\epsilon} \right)^{(L-1)/2} e^{-(x + L\epsilon)} I_{L-1}(2\sqrt{L\epsilon}x) dx \quad (2.28)$$

where ϵ is the signal-to-interference ratio (SIR) defined as: $\epsilon = \frac{NE_c R^2(\tau)}{V_o}$.

The correct and false detection probabilities in (2.27) and (2.28) are valid for additive white Gaussian noise (AWGN) channels, since we have assumed the received chip energy E_c is constant. In the Rayleigh fading channel, E_c should be multiplied by α^2 , where α is a Rayleigh-distributed random variable with pdf given by [19]:

$$p(\alpha) = \frac{2\alpha}{\sigma^2} e^{-\alpha^2/\sigma^2} \quad (2.29)$$

where σ^2 is the second moment of α .

The expression for P_F has the same form as (2.27) in the Rayleigh fading channel with the exception that V_o is replaced with $\bar{V}_o = N_0 + (2/3) \sum_{j \neq k} \sigma^2 E_c(j)$ (see (2.19)). The correct detection probability is derived as follows. First, $p_1(Z)$ in (2.24a) in the Rayleigh fading channel should be like this:

$$p_1(Z) = \int_0^\infty p_1(Z|\alpha) p(\alpha) d\alpha \quad (2.30)$$

Since: $I_0(t) = \frac{1}{2\pi} \int_0^{2\pi} \exp(t \cos \phi) d\phi$, (2.30) can be written as:

$$p_1(Z) = \int_0^\infty \frac{2\alpha}{\sigma^2} e^{-\frac{\alpha^2}{\sigma^2}} \frac{1}{NV_o} e^{-\frac{Z + \alpha^2 N^2 E_c R^2(\tau)}{NV_o}} d\alpha \int_0^{2\pi} \frac{1}{2\pi} e^{\frac{2\alpha \sqrt{ZE_c} \cos \phi}{V_o}} d\phi \quad (2.31a)$$

If we make the following variable changes:

$$x = \alpha \cos \phi$$

$$y = \alpha \sin \phi$$

$$dx dy = \alpha d\alpha d\phi$$

(2.31a) takes on the following form:

$$\begin{aligned} p_1(Z) &= \frac{1}{\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{1}{\sigma^2 NV_o} e^{-\frac{x^2 + y^2}{\sigma^2}} e^{-\frac{Z + N^2 E_c R^2(\tau)(x^2 + y^2)}{NV_o}} e^{\frac{2xR(\tau)\sqrt{ZE_c}}{V_o}} dx dy \\ &= \frac{1}{\pi \sigma^2 NV_o} e^{-\frac{Z}{NV_o}} \int_{-\infty}^{\infty} e^{-\frac{x^2}{\sigma^2} - \frac{NE_c R^2(\tau)x^2}{V_o} + \frac{2xR(\tau)\sqrt{ZE_c}}{V_o}} dx \int_{-\infty}^{\infty} e^{-\frac{y^2}{\sigma^2} - \frac{NE_c R^2(\tau)y^2}{V_o}} dy \\ &= \frac{1}{NV_f} e^{-Z/NV_f} \end{aligned} \quad (2.31b)$$

where $V_f = V_o + N\sigma^2 E_c R^2(\tau) = V_o + NE[\alpha^2] E_c R^2(\tau) = V_o(1 + \bar{\epsilon})$, and $\bar{\epsilon} = E[\alpha^2] \epsilon$.

An examination of (2.31b) and (2.23a) reveals that (2.31b) is virtually the same as (2.23a) except for the scaling factor V_f . Hence, the probability of correct detection in the Rayleigh fading channel has the same form as (2.27), but with V_o replaced by V_f .

$$P_D = e^{-\theta/NV_f} \sum_{k=0}^{L-1} \frac{(\theta/NV_f)^k}{k!} \quad (2.32)$$

Probabilities of both correct and false acquisitions for QPSK spreading are the same as those for BPSK [17].

To evaluate the synchronizer's performance, the following parameter values are assumed.

- Chip rate 1.2288 Mcps.
- Matched filter lengths 128.
- Synchronization message consisting of several repetitions of a 128 chip PN sequence.

The design objective is to keep the probability of false acquisition below 10^{-6} . To achieve that, we need to find the normalized threshold θ from (2.27) first. We can then compute the probability of correct acquisition from equations (2.28) and (2.32) for the Gaussian and Rayleigh fading channels, respectively. The probability of correct acquisition versus SIR for both the Gaussian and Rayleigh fading channels is plotted in Fig. 2.17 and Fig. 2.18 when the probability of false acquisition is specified to be 10^{-6} . It is seen that, in a Rayleigh fading channel, when the SIR is around 12 dB the synchronizer acquires synchronism over one preamble containing 32 repetitions of 128 chip PN sequence, which is equal to 3.3 ms at the chip rate of 1.2288 Mchips/s.

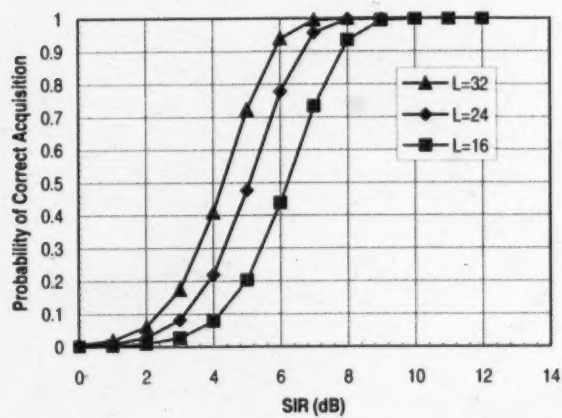


Fig. 2.17. Probability of correct acquisition in the Gaussian channel.

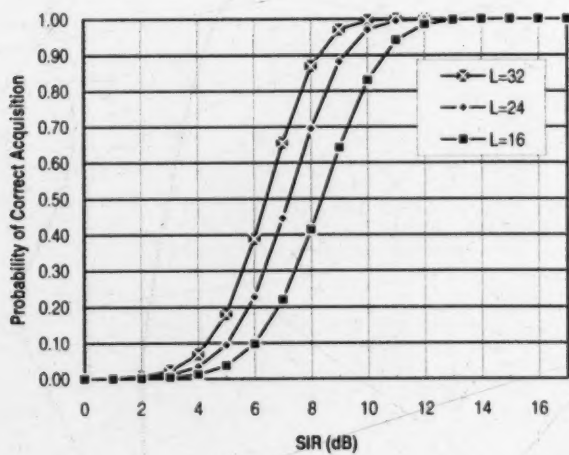


Fig. 2.18. Probability of correct acquisition in the Rayleigh fading channel.

2.5 Delay/throughput analysis

2.5.1 SDTX-MAC Scheme.

To gain an insight into the influence of false acquisition events on the overall data

transmission delay, consider initially a simplified case when all access request messages within a sequence are transmitted at the same power level (which implies that there are no distinguishable ARM sequences). Suppose the synchronization is acquired at the $(i+1)$ -th frame, and there have been k false acquisitions. The random access delay is then:

$$T(i, k) = (i+1)M_F T_p + kT_{fa} \quad (2.33)$$

where M_F is the number of slots in one ARC frame, T_p is the slot length, and T_{fa} is the delay penalty when a false acquisition occurs. The probability of the delay $T(i, k)$ is:

$$P(i, k) = P_D (1 - P_D)^i \binom{i}{k} P_F^k (1 - P_F)^{i-k} \quad (2.34)$$

where P_D and P_F are the correct and false acquisition probabilities of the synchronizer.

Combining equations (2.33) and (2.34), we obtain the mean access delay as follows:

$$\begin{aligned} E(T) &= \sum_{i=0}^{\infty} \sum_{k=0}^i T(i, k) P(i, k) \\ &= \sum_{i=0}^{\infty} \sum_{k=0}^i ((i+1)M_F T_p + kT_{fa}) P_D (1 - P_D)^i \binom{i}{k} P_F^k (1 - P_F)^{i-k} \\ &= \sum_{i=0}^{\infty} P_D (1 - P_D)^i \left[(i+1)M_F T_p \sum_{k=0}^i \binom{i}{k} P_F^k (1 - P_F)^{i-k} + T_{fa} \sum_{k=0}^i k \binom{i}{k} P_F^k (1 - P_F)^{i-k} \right] \\ &= \sum_{i=0}^{\infty} P_D (1 - P_D)^i [(i+1)M_F T_p + iP_F T_{fa}] \\ &= M_F T_p P_D \left[\sum_{i=0}^{\infty} i(1 - P_D)^i + \sum_{i=0}^{\infty} (1 - P_D)^i \right] + P_D P_F T_{fa} \sum_{i=0}^{\infty} i(1 - P_D)^i \\ &= (M_F T_p + P_F T_{fa}) \frac{1 - P_D}{P_D} + M_F T_p \end{aligned} \quad (2.35)$$

where the following identities have been used:

$$\sum_{k=0}^i \binom{i}{k} p^k (1-p)^{i-k} = 1; \quad p < 1$$

$$\sum_{k=0}^i k \binom{i}{k} p^k (1-p)^{i-k} = ip$$

$$\sum_{k=0}^{\infty} p^k = \frac{1}{1-p}$$

$$\sum_{k=0}^{\infty} kp^k = \frac{p}{(1-p)^2}$$

From Eqn. (2.35), we see that the effect of false acquisition penalty is negligible when P_F is very small. As explained in [20], the synchronizers are designed to keep the false acquisition probability P_F below 10^{-6} . Therefore, even if the false penalty delay T_{fa} is thousand times larger than T_P , $P_F T_{fa}$ is still trivial compared with $M_F T_P$. Hence, we will not consider the effect of false acquisition events in the following.

Now consider the case when the ARMs within a sequence have increasing power levels, and suppose that synchronism is acquired at the $(n+1)$ -th ARM of the $(j+1)$ -th ARM sequence. The probability of such event is:

$$P(n, j) = \prod_{i=1}^{M_s} (1 - P_{Di})^j \prod_{i=0}^n (1 - P_{Di}) P_{D(n+1)}; \quad P_{D0} = 0 \quad (2.36)$$

where P_{Di} is the i -th ARM's detection probability, and M_s is the number of ARMs in one ARM sequence. The random access delay is then:

$$T(n, j) = jM_s T_F + (n+1)T_F; \quad T_F = M_F T_P \quad (2.37)$$

Combining (2.36) and (2.37) we obtain the following equation for the mean access delay

of the SDTX-MAC scheme:

$$E(T) = \sum_{j=0}^{\infty} \sum_{n=0}^{M_s-1} T(n, j) P(n, j) \quad (2.38)$$

To compute throughput efficiency, we first need to consider how many ARMs need to be transmitted to gain access. That number is given by:

$$N_s = \sum_{j=0}^{\infty} \sum_{k=1}^{M_s} (jM_s + k) \left[\prod_{k=1}^{M_s} (1 - P_{Dk}) \right]^j \prod_{i=1}^{k-1} (1 - P_{Di}) P_{Dk} \quad (2.39a)$$

If $P_{D1} = P_{D2} = \dots = P_{DM_s} = P_D$, (2.39a) can be readily simplified to become:

$$N_s = 1 / P_D \quad (2.39b)$$

Eqn. (2.39b) is consistent with (2.35) for the case of negligible false acquisition events.

The throughput efficiency, or the ratio of the number of ARMs successfully acquired by the receiver to the number of transmitted ARMs can then be expressed as:

$$\eta = 1 / N_s \quad (2.40)$$

2.5.2 ADTX-MAC Scheme.

The ADTX-MAC scheme is in principle a spread slotted ALOHA as explained in section 2.2.2. Before embarking on analysis of this scheme, let us introduce the basics of the original ALOHA systems [4], [21]. There are two versions of conventional ALOHA systems: pure or unsynchronized and slotted or synchronized. In a pure ALOHA system, a user may begin transmitting a packet at any time. In a slotted ALOHA system, a user can transmit a packet at the beginning of a time slot only. A simple example (two users

only) of the operation of the ALOHA system for both pure and slotted versions is shown in Fig. 2.19 and Fig. 2.20. Note that in the conventional ALOHA system (pure or slotted), packets are not spread. Hence, whenever there is a time overlap, collision occurs, and the packets are destroyed. If we assume that start times of transmitted packets form a Poisson point process with the arrival rate λ packets/s, and denote the duration of a packet in seconds as T_p , then the offered channel traffic G will be

$$G = \lambda T_p \quad (2.41)$$

In a pure ALOHA system, collision is avoided if no packets arrive T_p before or T_p after the arrival time of the given packet. Since the start times of all packets are Poisson-distributed, the probability that k packets are generated during a given time t is given by [22]:

$$P(k) = \frac{(\lambda t)^k e^{-\lambda t}}{k!} \quad (2.42)$$

The probability of no collision, which means that zero packets are generated within the time duration of $2T_p$, is $P(0) = e^{-2\lambda T_p} = e^{-2G}$. Therefore the pure ALOHA channel throughput is:

$$S = G e^{-2G} \quad (2.43)$$

In a slotted ALOHA system, since a packet can only be transmitted at the beginning of each time slot, the probability of no collision is thus the probability of an event that no packets are generated during the period T_p . The probability of such an event is $e^{-\lambda T_p}$, or e^{-G} . The throughput of a slotted ALOHA channel is thus given by:

$$S = Ge^{-G} \quad (2.44)$$

Fig. 2.21 shows the throughput characteristics for pure and slotted ALOHA

The maximum throughput for the pure ALOHA can be found by first taking the derivative of equation (2.43) and equating it to zero to obtain the value of G ,

$$\frac{dS}{dG} = e^{-2G} - 2Ge^{-2G} = 0 \Rightarrow G = \frac{1}{2} \quad (2.45)$$

and substituting this value into (2.43). The maximum throughput is found to be $1/2e = 0.184$, which can also be read from Fig. 2.21. Using the same method, the maximum throughput for the slotted ALOHA is $1/e = 0.368$ when the offered traffic is 1, twice the capacity of the pure ALOHA channel.

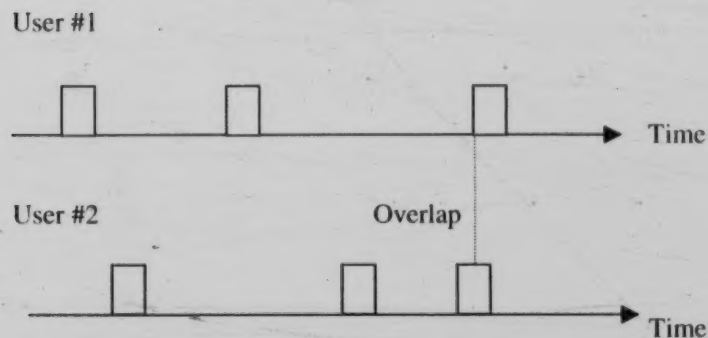


Fig. 2.19. Packets from two users on a pure ALOHA channel.

A p -persistent slotted ALOHA is designed to reduce the rate of collisions which in principle is similar to the p -persistent carrier sense multiple access (CSMA) protocol [7].

The p -persistent slotted ALOHA can be simply described as follows. Before transmitting each packet other than the first packet, the mobile station shall generate a random number in the range $<0,1>$, and if this number is larger than the probability p , the mobile station shall delay transmission of the packet for a pre-determined number of slots.

User #1



Fig. 2.20. Packets from two users on a slotted ALOHA channel.

The ADTX-MAC scheme is p -persistent spread slotted ALOHA. Due to the slotted ALOHA structure, there is certain probability that one ARM collides with one or more ARMs from other users. However, the collision event in the spread ALOHA is different from that in the conventional ALOHA. In the conventional ALOHA system, any packet overlap implies collision. In the spread ALOHA system, due to the auto-correlation property of PN sequences, time overlap of two packets may not cause collision. The autocorrelation function for a maximum-length PN sequence with chip duration T_c and period NT_c [15] is illustrated in Fig. 2.22. Normally, a PN sequence with a very long period is used for spreading. For example in the IS-95 system [3], the length or period of

a mobile-specific PN sequence is $N = 2^{42} - 1$; and $\frac{1}{N}$ is about 2.3×10^{-13} . Hence, if the separation time between two mobiles using the same spreading sequence is more than one chip (T_c), one appears like noise to another.

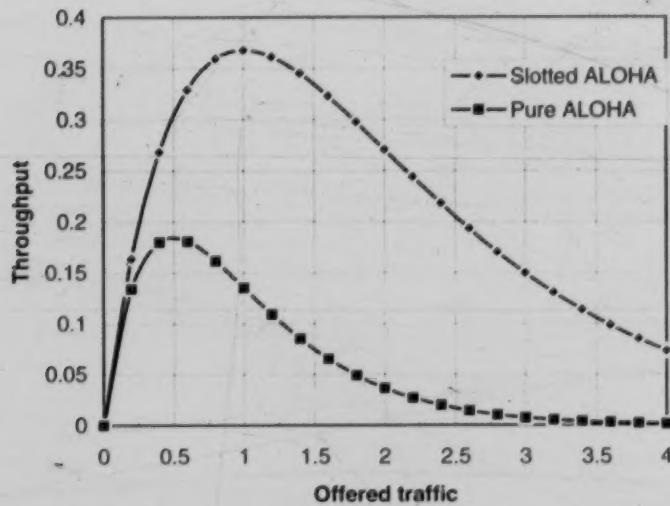


Fig. 2.21. Throughput versus offered traffic for the pure and slotted ALOHA.

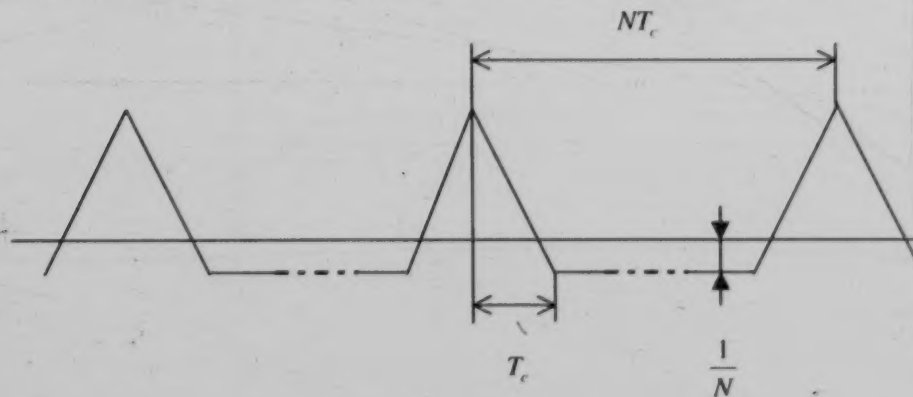


Fig. 2.22. Autocorrelation function for a maximum-length PN sequence.

In general, for the ADTX-MAC scheme in flat fading channels the collision occurs only when the relative delay of arriving ARMs is less than one chip time. However, when a transmitted signal propagates through the multipath fading channel, its replicas will arrive at the receiving end at different times spread over the maximum excess delay interval. If we denote the maximum excess delay expressed in chip intervals by τ_M , it is reasonable to assume that when the temporal separation of ARMs arriving from different users is larger than $(\tau_M + 1)$ chip intervals, collisions will not occur. The precise timing of the ARM transmissions in the ADTX-MAC protocol is determined by a procedure called PN randomization. Before transmitting any ARM, the mobile station computes a delay which depends on its electronic serial number (ESN), and delays its transmission by the computed value. Suppose that the maximum PN randomization delay is R_M . The probability that a given user's ARM will not collide with other ARMs is the probability that no ARM starts $(\tau_M + 1)$ chips before or $(\tau_M + 1)$ chips after the start time of the given ARM. Since each ARM's arrival time is uniformly distributed on $(0, R_M)$, the probability of no collision conditioned on the event that $k+1$ ARMs (one desired and k interfering ARMs) are sent within the interval $(0, R_M)$ is:

$$P(\text{no\_collision} | k) = \begin{cases} 1 & k = 0 \\ (1 - 2\psi)^k & k \geq 1 \end{cases} \quad (2.46)$$

where $\psi = (\tau_M + 1) / R_M$. Lets us assume an infinite population model for all bursty data users, with the offered traffic of ARMs of G messages per slot. The probability that there are k ARMs transmitted within one slot is Poisson distributed:

$$P(k) = \frac{G^k e^{-G}}{k!} \quad (2.47)$$

The probability that an ARM will not collide with any other ARM is:

$$P(\text{no\_collision}) = \sum_{k=0}^{\infty} P(\text{no\_collision} | k) P(k) = e^{-2G\psi} \quad (2.48)$$

Assuming that the synchronizer is perfect ($P_D = 1, P_F = 0$), the throughput of ADTX-MAC denoted as S_1 is then given by:

$$S_1 = G e^{-2G\psi} \quad (2.49)$$

The throughput versus offered traffic is plotted in Fig. 2.23.

It can be readily shown that the throughput reaches its maximum at $S_2 = 1/2e\psi$ when offered traffic is equal to $1/2\psi$. If we increase the offered traffic above $1/2\psi$, the throughput will decrease. It can also be seen from the figure that the longer the maximum randomization delay, the larger the maximum throughput, but naturally the longer the access delay.

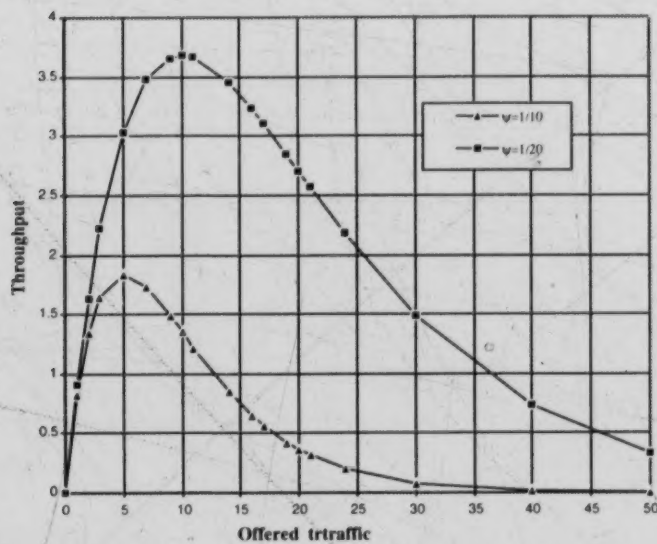


Fig. 2.23. Throughput versus offered traffic for a spread ALOHA channel.

Theoretically, in an interference-limited spread spectrum system, there is a maximum allowed number of users in the channel at the same time. If the simultaneous transmitters exceed that maximum, the system's performance will degrade dramatically. In general, the throughput can be expressed as:

$$S_2 = G \sum_{k=0}^K P(\text{no\_collision} | k) P(k) \quad (2.50)$$

where K is finite, whereas (2.49) was obtained with no limit on the maximum value. However, to simplify our analysis, (2.49) is used in the remainder of this paper.

Suppose that the access request is correctly detected at the $(n+1)$ -th ARM of the $(j+1)$ -th ARM sequence. The probability of such an event is:

$$P(n, j) = \prod_{i=1}^{M_r} (1 - P_{Di} P_{Ci})^j \prod_{i=0}^n (1 - P_{Di} P_{Ci}) P_{D(n+1)C(n+1)}; P_{D0} = P_{C0} = 0 \quad (2.51)$$

where P_{Ci} is the probability that no error is detected within the appendix of the i -th ARM, which is the function of the signal to noise ratio (SNR); all other symbols have the same meaning as in (2.36). The probability that no error is detected within the appendix actually consists of P_{ne} , the probability that a received message contains no error, plus P_{ue} , the probability that the received appendix contains an undetectable error pattern.

For a primitive (127, 64) BCH code with minimum distance 21, it can be readily shown that P_{ue} is much smaller than P_{ne} . Therefore, $P_{Ci} \approx P_{ne}$ can be assumed. The average probability that a received message contains no error is given by:

$$P_{ne} = (1 - P_b)^{96} \quad (2.52)$$

where P_b is the average bit error probability of an optimal noncoherent receiver for M -ary orthogonal signaling on a Rayleigh fading channel, which is given by [19] as follows:

$$P_b = \frac{M/2}{M-1} \left(1 - \int_0^\infty \frac{1}{(L-1)! \left(1 + \frac{\log_2 M}{L} \gamma\right)^L} x^{L-1} \exp\left(-\frac{x}{1 + \frac{\log_2 M}{L} \gamma}\right) \left(1 - e^{-x} \sum_{k=0}^{L-1} \frac{x^k}{k!}\right)^{M-1} dx \right) \quad (2.53)$$

where γ is the average SNR per bit, and L is the number of diversity branches. Unfortunately, (2.53) is computationally cumbersome to evaluate for large values of M and L . Alternatively, for our later numerical calculations, we may use its upper bound which is also given by [19]:

$$P_b \leq \frac{M}{2} p_o^L \sum_{k=0}^{L-1} \binom{L-1+k}{k} (1 - p_o)^k \quad (2.54)$$

where $p_o = 1/(2 + \gamma(\log_2 M)/L)$ is the average error probability for binary orthogonal signaling on a Rayleigh fading channel without diversity.

The delay corresponding to acquisition in the $(n+1)$ -th ARM of the $(j+1)$ -th ARM sequence is:

$$T(n, j) = (n + jM_s)(T_A + T_R + T_M + \hat{R}_M) + j(T_s + T_{PD}) + T_M + \hat{R}_M \quad (2.55)$$

where T_A is the acknowledgment time-out, T_R is the ARM random back-off delay, T_M is the access request message length, $\hat{R}_M$ is the randomization delay introduced before any ARM transmission (it depends on the mobile's ESN), T_s is the ARM sequence random back-off delay, and T_{PD} is the p -persistence delay introduced before each new ARM sequence. T_{PD} depends on the parameter p of the p -persistence test, and its average value is calculated as follows:

$$\bar{T}_{PD} = T_M \sum_{i=0}^{\infty} i(1-p)^i p = \frac{1-p}{p} T_M \quad (2.56)$$

Consequently, the mean access delay for the ADTX-MAC scheme can be evaluated as:

$$E(T) = \sum_{j=0}^{\infty} \sum_{n=0}^{M_s-1} P(n, j) T(n, j) \quad (2.57)$$

Just as in the SDTX-MAC scheme, to compute the throughput we first evaluate the number of ARMs that must be transmitted to achieve successful acquisition:

$$N_A = \sum_{j=0}^{\infty} \sum_{k=1}^{M_s} (jM_s + k) \left[\prod_{k=1}^{M_s} (1 - P_{Dk} P_{Ck}) \right]^j \prod_{i=1}^{k-1} (1 - P_{Di} P_{Ci}) P_{Dk} P_{Ck} \quad (2.58)$$

Then, the throughput can be expressed as:

$$S_1 = \frac{Ge^{-2G\psi}}{N_A} \quad (2.59)$$

(2.59) is more general than (2.49), because it allows for $P_{Dk} < 1$ and $P_{Ck} < 1$. With $P_{Dk} = P_{Ck} = 1$, (2.59) becomes identical to (2.49).

2.5.3 HDTX-MAC Scheme.

The HDTX-MAC protocol's structure and parameters are presented in Fig. 2.4. As discussed earlier, clearly if T_m (the length of the minislot) is larger than $(\tau_M + 1)$, more than one ARM can be transmitted within one regular time slot. Assume the number of minislots in one slot is N (Fig. 2.4), and all ARMs are equal in length. Then

$$T_{LP} = (N - 1)T_m + T_p \quad (2.60)$$

Note that T_m is much smaller than T_p , the regular slot duration in the SDTX scheme, and N is mainly limited by the spread spectrum system's processing gain. For example, for an outdoor urban microcellular channel, τ_M is around 13 chips (at a chip rate of 1.2288 Mchips per second)[23], and N could be around $0.1P_G$ to $0.2P_G$ [5], where P_G is the spread spectrum system's processing gain. The message length T_p is 4096 chips if the access preamble consists of 32 repetitions of 128 PN chips. Thus, T_{LP} is only moderately longer than T_p of the SDTX slot.

The mean access delay of the HDTX-MAC scheme is still given by (2.38), with T_p

replaced by T_{LP} , but the correct acquisition probability P_d which is implicitly included in $P(n, j)$, is different from the one for SDTX-MAC due to the multi-user interference in the HDTX-MAC scheme. Assuming the signal-to-noise ratio (SNR) in the SDTX-MAC scheme is ϵ , with the same transmitting power and background noise, the SNR in the HDTX-MAC scheme is $\frac{1}{1/\epsilon + \alpha}$ where $N-1 = \alpha P_G$. For example, if the SNR in SDTX-MAC is 10 dB, then the SNR in HDTX-MAC is about 7 dB when $\alpha = 0.1$. Hence, the mean access delay of the HDTX-MAC is longer than that of the SDTX-MAC scheme. However, the throughput of the HDTX-MAC is orders of magnitude higher than that of the SDTX-MAC, since in the SDTX-MAC scheme only one user is allowed to transmit his ARM within one time slot, and many users may do that under HDTX-MAC.

2.6 Performance evaluation

It is observed from the above analysis that the delay distributions of the two protocols to a large degree depend on P_d , the correct detection probability of the access request message. To obtain numerical results illustrating the performance of the three protocols, the following assumptions have been made:

- Chip rate is 1.2288 Mchips/s.
- A microcell CDMA network is considered with radius 1 km; therefore the maximum round trip delay is approximately 8 chips.
- The Power increment between ARMs is 2 dB.
- ARM length in SDTX-MAC contains 32 repetitions of the 128 chip PN sequence.

- The maximum number of ARMs within one sequence is $M_s = 16$.
- The maximum number of ARM sequences in one acquisition attempt is 15.
- The Doppler shift is 100 Hz.
- There are 25 slots in each access frame on the ARC for SDTX-MAC.
- The value of p for the p -persistence test is 0.5.
- The maximum randomization delay R_M is 280 chips.
- T_s , the sequence back-off range is 0~4 slots.
- T_R , the ARM back-off range is 0~3 slots.
- T_A , the ACK response time-out, is equal to the maximum round trip delay plus the maximum search and verification time of the synchronizer.
- No diversity is considered ($L=1$).
- The probability of correct acquisition is evaluated in the presence of Rayleigh fading.

As mentioned before, for the urban outdoor microcellular channel assumed here, the maximum excess delay τ_M at the chip rate of 1.2288 Mcips/sec is 13 chips [23]. That value of τ_M corresponds to the 22 dB drop of the multipath intensity profile. Therefore $\psi = 1/20$ is used in the numerical analysis.

The mean access delays for both SDTX-MAC and ADTX-MAC protocols are shown in Fig. 2.24 and Fig. 2.25. From Fig. 2.25 it is observed that at low SNR, the mean access delay of ADTX-MAC is much larger. The reason is that, at the receiving end, the probability of a bit error of the ID message is high, which naturally leads to more

retransmissions, and therefore longer access delay. The mean access delay for errorless ID transmission is shown in Fig. 2.26. In that case the access delay depends only on the performance of the dedicated synchronizer. The throughput for SDTX-MAC protocol is shown in Fig. 2.27. Note that the throughput for the SDTX-MAC may approach one, but can not exceed one even for an arbitrarily high signal to noise ratio. For the ADTX-MAC protocol, beyond a threshold value of the offered traffic which depends on the system parameters, the probability of collisions of access request messages from different users is very high. Fig. 2.28 shows the throughput curves for the ADTX-MAC scheme as a function of offered traffic at different signal to interference ratios of the initial ARM. It is seen that the ADTX-MAC scheme's throughput can be more than one per slot depending on the traffic, and the throughput increases approximately linearly before reaching its peak value.

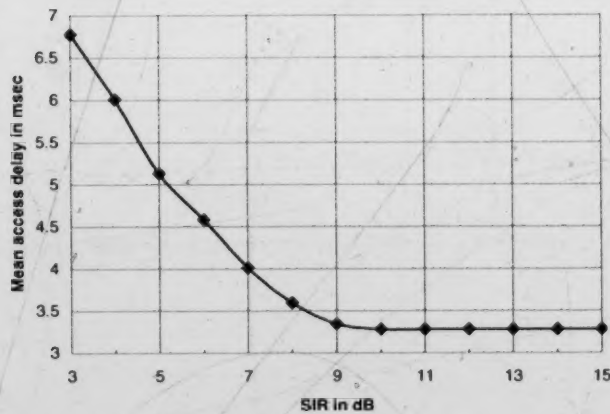


Fig. 2.24. Mean access delay versus the initial ARM's SNR for the SDTX-MAC protocol.

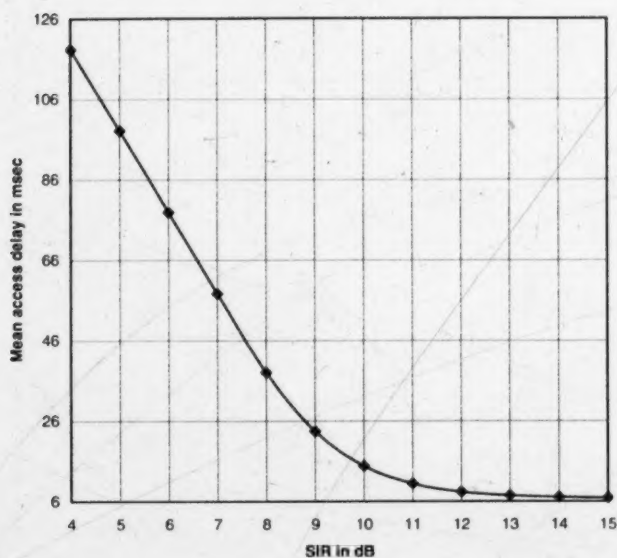


Fig. 2.25. Mean access delay versus the initial ARM's SNR for the ADTX-MAC.

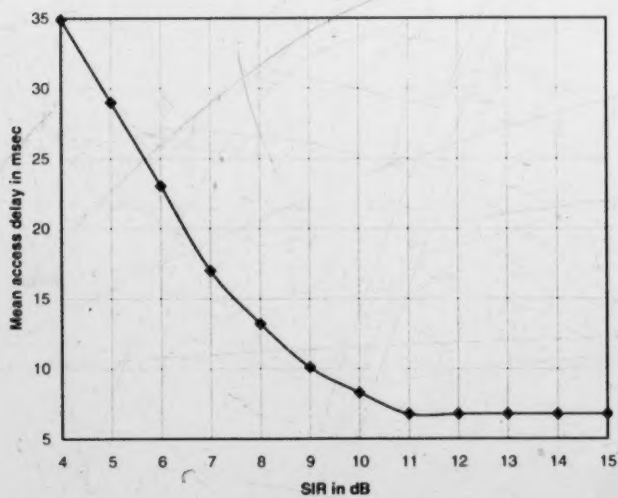


Fig. 2.26. Mean access delay versus the initial ARM's SNR for ADTX-MAC protocol when the ID transmission is errorless.

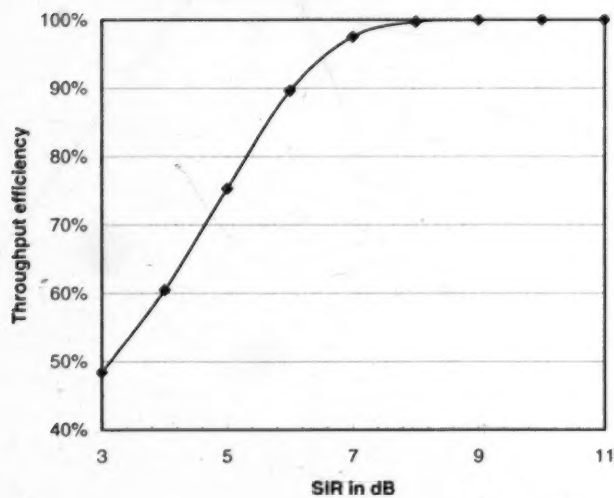


Fig. 2.27. Throughput efficiency versus the initial ARM's SNR for SDTX-MAC .

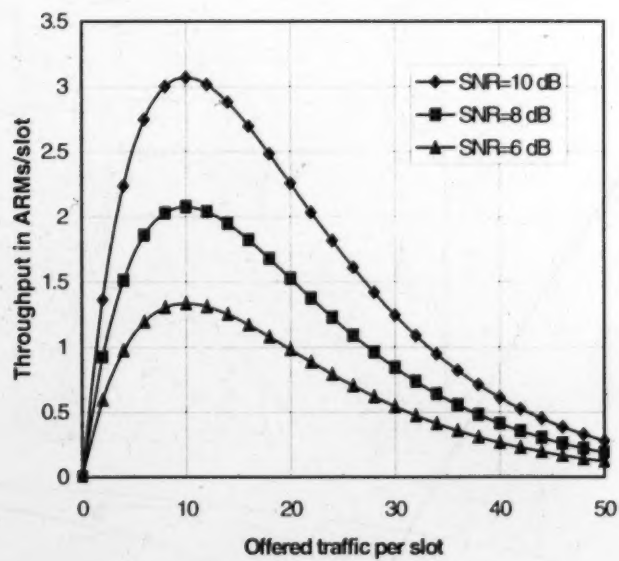


Fig. 2.28. Throughput versus offered traffic for the ADTX-MAC protocol.

The throughput for the HDTX-MAC scheme in ARMs per slot is around $0.1P_G$ to $0.2P_G$ at high SNRs. The curve of the mean access time for the HDTX-MAC scheme versus SNR is similar to that of the SDTX-MAC shown in Fig. 2.24. However, if the required SNR is the same for both schemes, the HDTX-MAC scheme needs more transmitting power than the SDTX-MAC does.

2.7 Conclusions and discussion

We have demonstrated that the synchronous discontinuous transmission MAC (SDTX-MAC) protocol can provide delay characteristics superior to these of the asynchronous scheme (ADTX-MAC). One of the reasons is that ADTX-MAC is inherently unstable due to the possibility of collisions, and another is that ADTX-MAC's access request message must include a field indicating the identity of the mobile, for which the probability of error must be very low. The SDTX-MAC eliminates collisions between ARMs sent by different users, and does not require the mobile's identity to be transmitted in the ARM. Hence, the length of an ARM in SDTX-MAC is less than half of that in ADTX-MAC, and therefore its access delay is lower. As far as the throughput is concerned, when the offered traffic is low, ADTX-MAC may provide higher throughput than the SDTX-MAC, but when the offered traffic is high, the chance for an ARM being captured by the receiver using ADTX-MAC is slim. The proposed hybrid synchronous MAC combines advantages of both the SDTX and ADTX-MAC schemes. Its mean access delay depends on the number of minislots (or users) within each regular slot, but its throughput is much higher than that of either SDTX- or ADTX-MAC. For wideband

CDMA systems which can support large number of users, the HDTX-MAC is preferred.

The idea behind the HDTX-MAC relies on the fact that different codes in a CDMA system are not necessary to distinguish among different transmitters [5], [6]. The HDTX-MAC scheme not only takes advantage of the wideband nature of spread spectrum, but also solves the potential collision problem of the spread ALOHA multiple access (SAMA) method proposed in [5] and [6]. HDTX-MAC scheme constitutes a spread time division structure. Hence, the idea of the HDTX-MAC scheme is not limited to the application in medium access control. It can be used as a general multiple access method especially in CDMA/TDMA co-existence systems. The distinct feature of the hybrid synchronous multiple access scheme is that all transmitters are assigned the same spreading code, and therefore only a single matched filter or correlator is required in the receiver in order to receive all signals as explained in [5] and [6]. This feature greatly reduces the design complexity of base stations. On the other hand, system capacity with the hybrid synchronous multiple access method is comparable to that of the conventional CDMA system using a distinct spreading code for each user.

References

- [1] TR45.5.0.18, "The CDMA2000 ITU-R RTT Candidate Submission", June 1998.
- [2] A. Baier, "Open multi-rate radio interface architecture based on CDMA," in *Proc. ICUPC'93*, Ottawa, Canada, Oct. 1993, pp. 985-989.
- [3] TIA/EIA IS-95, "Mobile station - base station compatibility standard for dual-mode

wideband spread spectrum cellular systems," *Telecommunications Industry Association*, 1993.

[4] N. Abramson, "The throughput of packet broadcasting channels", *IEEE Trans. Communications*, vol. COM-25, no.1, January, 1977, pp. 117-128.

[5] N. Abramson, "Wideband random access for the last mile," *IEEE Personal Communications Magazine*, December 1996, pp. 29-33.

[6] N. Abramson, *Spread ALOHA CDMA Data Communications*, US Patent 5,537,397, July 16 1996.

[7] L. Kleinrock and F. A. Tobagi, "Packet switching in radio channels: Part I-Carrier sense multiple-access modes and their throughput-delay characteristics", *IEEE Trans. Communications*, COM-23, December 1975, pp. 1400-1416.

[8] S.B. Wicker, *Error Control Systems for Digital Communication and Storage*, Prentice-Hall, New Jersey, 1995.

[9] M. K. Simon, J. K. Omura, R. A. Scholtz, and B. K. Levitt, *Spread Spectrum Communications Handbooks*, revised edition, McGraw-Hill Inc., 1994.

[10] J. K. Holmes and C. C. Chen, "Acquisition time performance of PN spread spectrum systems", *IEEE Trans. Commun.*, COM-25, August 1977, pp. 778-783.

[11] A. Polydoros and C. L. Weber, "A unified approach to serial search spread spectrum code acquisition: Part I. General theory", *IEEE Trans. Commun.*, Vol. COM-32, May 1984, pp. 542-549.

[12] D. M. DiCarlo and C. L. Weber, "Statistical performance of single dwell serial synchronization systems", *IEEE Trans. Commun.*, COM-28, no. 8, August 1980, pp.

1382-1388.

[13] D. M. DiCarlo, *Multiple dwell serial synchronization of pseudonoise signals*, Ph.D. Dissertation, Department of Electrical Engineering, University of Southern California, May 1979.

[14] D. M. DiCarlo and C. L. Weber, "Multiple dwell serial search: Performance and application to direct sequence code acquisition", *IEEE Trans. Commun.*, COM-31, no. 5, May 1983, pp. 650-659.

[15] R.L. Peterson, R.E. Ziemer and D.E. Borth, *Introduction to Spread Spectrum Communications*, Prentice-Hall, New Jersey, 1995.

[16] A. Polydoros and C. L. Weber, "A unified approach to serial search spread spectrum code acquisition: Part II. A matched filter receiver", *IEEE Trans. Commun.*, Vol. COM-32, May 1984, pp. 550-560.

[17] A. J. Viterbi, *CDMA: Principles of Spread Spectrum Communication*, Addison-Wesley, 1995.

[18] W.-C. Lin, K.-C. Liu and C.-K. Wang, "Differential matched filter architecture for spread spectrum communication systems," *Electron. Letters*, Vol.32 No.17, August 1996.

[19] J.G. Proakis, *Digital Communications*, 3rd edition, McGraw-Hill, New York, 1995.

[20] S. Sun and W. A. Krzymien, "Matched filter based synchronizer design for rapid synchronization of packet data transmissions in DS-CDMA personal communications systems," in *Proc. PACRIM'97*, Victoria, Canada, Aug. 1997, pp. 5-8.

[21] L. G. Roberts, "ALOHA packet system with and without slots and capture", *Comput. Commu. Rev.*, vol. 5, pp. 28-42, April 1975.

[22] A. Leon-Garcia, *Probability and Random Processes for Electrical Engineering*, Addison-Wesley, Reading, Mass., 1994.

[23] JTC(AIR)/93.09.23-238R2, "Report on RF channel characterization," Sept. 23, 1993, p. 20.

Chapter 3

DATA TRANSMISSION WITH HIGH RELIABILITY

3.1 Introduction

For highly reliable data transmission through radio channel, the forward error correcting (FEC) schemes, which are successfully used for voice communications, are not sufficient to meet the strict quality of service requirements for data. Existing computer data networks were designed to run in a wireline environment where the frame error rate (FER) is very small, normally below 10^{-9} . However the typical link FER in wireless cellular systems is around 1 percent. Network performance may suffer severe degradation if protocols designed for wireline networks are applied directly to the wireless cellular environment. Radio link protocols (RLP) are necessary in order to reduce the link frame error rate to the level of the wireline communication environment.

Highly reliable communication can be achieved using an automatic-repeat-request (ARQ) protocol. However, although a pure ARQ error control system in general can provide much more reliable service than an FEC system, it is inferior to the FEC in terms of throughput as the channel quality deteriorates. The drawbacks of the ARQ and FEC schemes can be overcome through the joint use of forward error correction and error detection in so-called hybrid ARQ protocols [1]-[2].

CDMA systems are interference limited. Powerful coding schemes have to be designed to correct as many errors as possible and therefore reduce the data retransmission rate in ARQ schemes. Convolutional codes have been proposed in IS-95 [3] and will be

employed here for error correction, not only because of their powerful error correction capabilities, but also relatively easy implementation of soft-decision decoding as compared with linear block codes.

To achieve high capacity and acceptable quality of service, the effects of fading and propagation loss are mitigated through the use of power control (both open and closed loop), which practically eliminates the near-far effect, and significantly reduces shadowing and multipath fading, especially at lower fading rates.

Cyclic codes are easily encoded and also allow for a number of convenient decoding techniques for error detection using shift register circuits. To achieve satisfactory quality of service, the codes need to have enough redundancy to guarantee negligible probability of undetected error. Unfortunately, some extensively used standard CRC codes, such as CRC-ANSI, CRC-CCITT, and IBM CRC-SDLC [2], [4], [5], [6] which have been used particularly in computer communications, are not designed for application in radio channels (and hence, on these channels standard CRC codes may not guarantee the desired very low probability of undetected error). For this reason, shortened cyclic codes, and more specifically, the shortened BCH codes are proposed for error detection in this work.

When an ARQ protocol is applied to data transmission, the retransmitted packets increase the interference level. If a call demanding high-reliability data service is accepted by the network, to account for retransmissions due to errors, the data subscriber's resultant

interference to the network should be calculated as the actual transmitted power times the average number of times that a packet must be transmitted before it is accepted. Clearly if the received signal-to-interference ratio is larger, the number of retransmissions is reduced; if the signal-to-interference ratio is smaller, the number of requested retransmissions increases. Clearly, there is an optimum signal-to-interference ratio which minimizes the cumulative interference and naturally leads to the increased capacity for an interference-limited CDMA system. Methods to find the optimum signal-to-interference ratio are presented in this thesis.

This chapter considers performance of a type I hybrid ARQ scheme (see Section 3.2.2.2 for the definition of type I hybrid ARQ) designed for high-reliability data transmission in DS-SS personal communication systems, and then determines the optimum signal-to-interference ratio for data transmission.

3.2 System description

3.2.1 Physical layer description

In a CDMA cellular system, each base station transmits a pilot signal on the forward link. All mobile stations listen to this pilot, and use it to obtain initial system synchronization and to provide time and phase tracking of the signals from the base station. However, on the reverse link, continuously transmitted pilot signals may not be available. Hence, the noncoherent detection technique is more applicable on the reverse link, although this may reduce the system capacity. Specifically a noncoherent link with M -ary orthogonal

modulation is considered here. The simplified block diagram of the proposed transceiver is shown in Fig. 3.1. In Fig. 3.1, the information bits are first encoded by a systematic BCH encoder, and then a convolutional encoder. The coded bits are grouped and mapped onto an M -ary orthogonal signal set, and block interleaved symbol-by-symbol. The signals prior to transmission are further spread on both in-phase and quadrature channels, and the data spread by the quadrature PN sequence is delayed by half a PN chip time with respect to the data spread by the in-phase PN sequence as presented in Fig. 3.2, where $a^I(t)$ and $a^Q(t)$ are the spreading waveforms of the in-phase and quadrature channels. Since the timing of the waveform streams of the in-phase and quadrature channels is shifted by half a chip, the carrier phase transitions of 180° are avoided. Hence, the waveform envelope will not go to zero, and the envelope variations are considerably reduced. This will permit more efficient RF amplification, which is particularly important for mobile communications on the reverse link where the battery life is a critical factor.

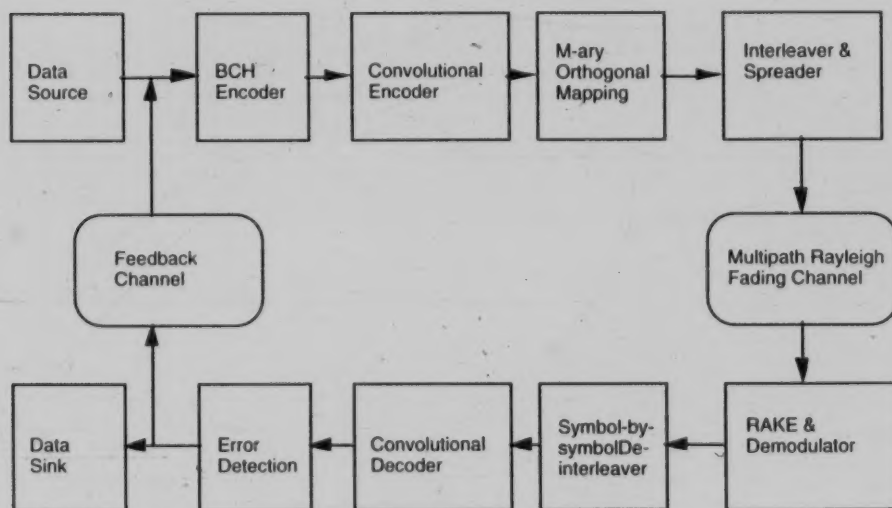


Fig. 3.1. Block diagram of a DS-CDMA data transmission system.

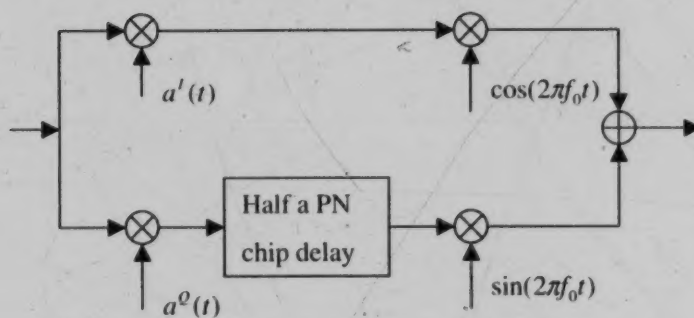


Fig. 3.2. Quadrature spreading of DS-CDMA signals.

A multipath fading phenomenon is caused by reflecting objects and scatters which due to the movement of the mobile station create a constantly changing dispersive propagation environment. Multiple versions of the transmitted signal may reach the receiving end in a multipath channel. The Rayleigh distribution is commonly used to describe the random nature of the received faded signal's envelope. A well-known structure used to track and receive several multipath signals is the so-called RAKE receiver [7]-[8]. The block diagram of a RAKE receiver is shown in Fig. 3.3. The receiver structure for one of the branches is shown in Fig. 3.4 [9]-[10]. In Fig. 3.3 and Fig. 3.4, we have assumed that there are L branches in the RAKE receiver. At the receiving end in Fig. 3.1, the received signals are first despread using an L -branch RAKE structure. M correlators are employed per branch to detect M -ary Walsh symbols. The M soft decisions (Walsh coefficients) are used for power control and in the Viterbi decoder. The output from the Viterbi decoder is fed into the error detector. If the decoded frame's syndrome is zero, the frame is declared error free, otherwise a retransmission is requested. In Fig. 3.4, $W'(t) (i=1,2,\dots,M)$ is the i -th orthogonal symbol waveform of the transmitted M -ary

signal. This orthogonal waveform is generated using the Walsh sequences or Hadamard codes. Hadamard codes can be generated by means of the following recursive procedure:

$$H_1 = 0; \quad H_2 = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}; \quad H_4 = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 \end{bmatrix}; \quad (3.1)$$

$$H_{2i} = \begin{bmatrix} H_i & \bar{H}_i \\ H_i & \bar{H}_i \end{bmatrix}$$

where i is a power of 2 and $\bar{H}_i$ denotes the binary complement of H_i . Note that if "0" and "1" in the Hadamard codes are mapped into "+1" and "-1", then the rows of the Hadamard codes are mutually orthogonal.

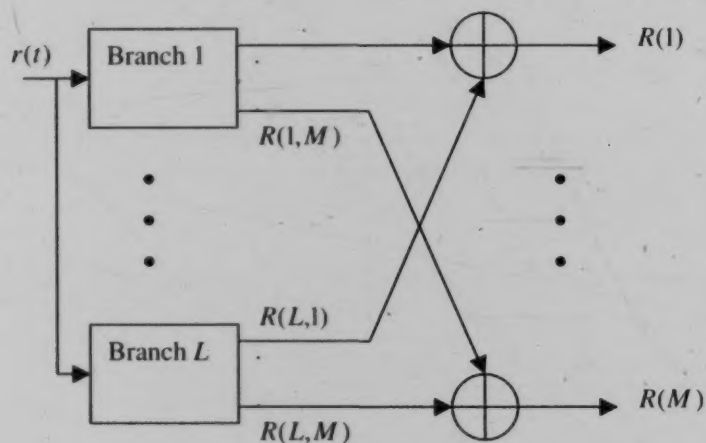


Fig. 3.3. Block diagram of the RAKE receiver.

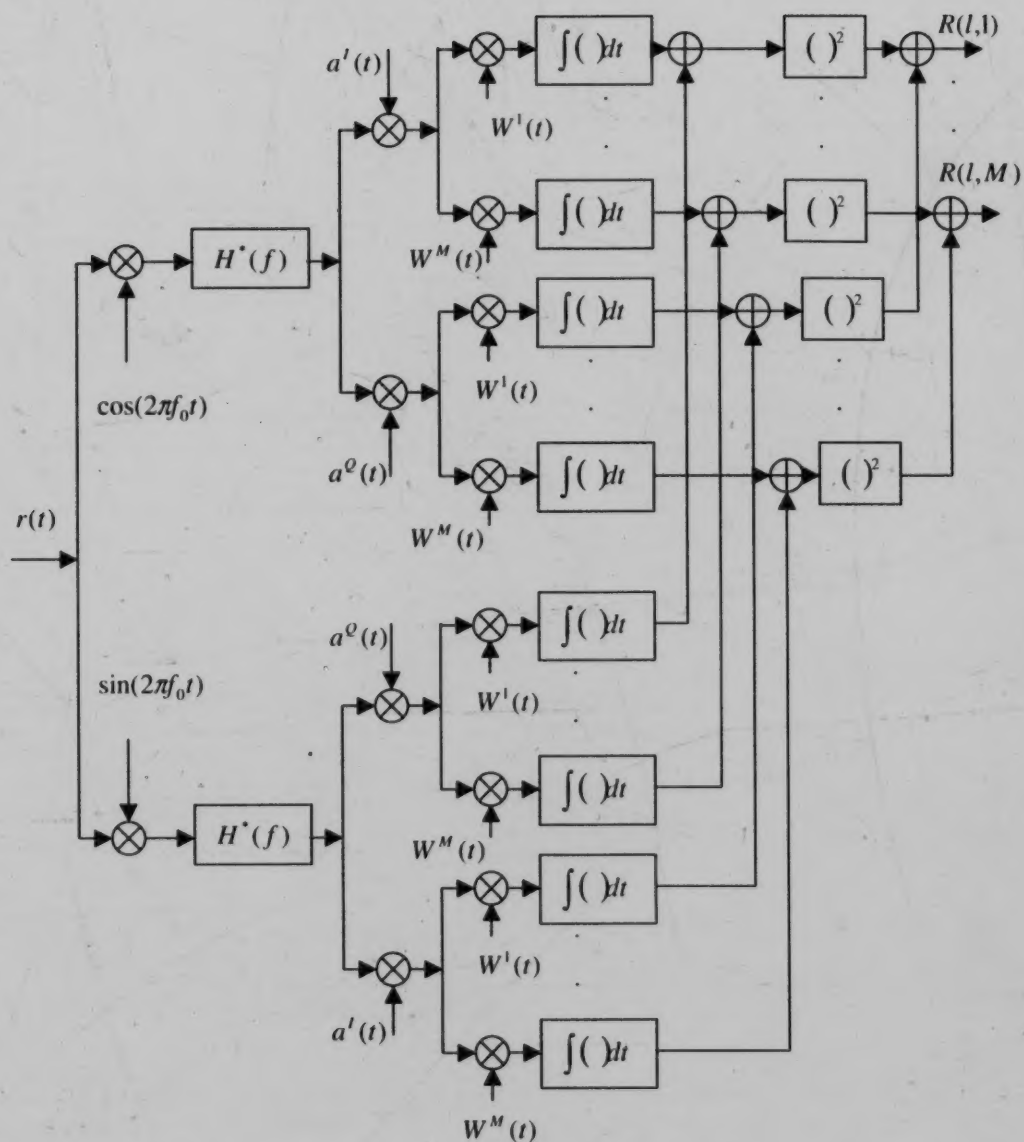


Fig. 3.4. The l -th branch of the RAKE receiver for M -ary orthogonal signaling.

Interleaving is a way of introducing temporal diversity into the signal. Normally, most channel codes are designed to combat random independent errors. However channel fades tend to cause errors in bursts. Interleaving the coded messages before transmission and

deinterleaving after reception breaks up bursts of errors and facilitates their correction. For the symbol-by-symbol block interleaver under consideration, a block of Walsh symbols is arranged as an $I \times J$ matrix. The Walsh symbols are written row by row and transmitted column by column. To combat the time-correlated impairments, the number of rows must be at least as large as the average fade duration. An ideal interleaver would be infinite in length, but it is not practically feasible because of its infinite latency. The number of columns J should be slightly larger than the number of coded symbols within a constraint length of the convolutional codes used. Thus, for a constraint length of 9, rate 1/3 convolutional coder and 64-ary orthogonal signaling, for the symbol-by-symbol interleaving case, a 96-Walsh symbol block can be arranged as an 16×6 matrix ($J = 6$, the code constraint length is 5 coded Walsh symbols). Note that here instead of using a conventional bit-by-bit interleaving technique, a symbol-by-symbol interleaving is employed. This is because there is no direct correspondence between demodulator envelopes and weight information per bit in the decoding process of the Viterbi algorithm (VA) for the convolutional coding combined with M -ary orthogonal signaling if a bit-by-bit interleaving is used. Thus, with a conventional bit-by-bit interleaving technique, the receiver may select the largest of the M outputs of the square-law multipath correlator and this value is used as a soft decision metric in the Viterbi decoding process [11]. In other words, some of the soft information is discarded when using bit-by-bit interleaving. All M outputs are used for decoding when a symbol-by-symbol interleaving is employed, as mentioned before. For this reason symbol-by-symbol interleaving performs better than bit-by-bit interleaving in this special case.

3.2.2 Retransmission protocols

3.2.2.1 Basic automatic-repeat-request (ARQ) strategies

Three basic retransmission protocols [1] are: the stop-and-wait (SW-ARQ), the go-back- N (GBN-ARQ), and the selective-repeat (SR-ARQ).

In a SW-ARQ system, the transmitter sends a data packet to the receiver and waits for an acknowledgment from the receiver as shown in Fig. 3.5. If the packet is received error free, a positive acknowledgment (ACK) will be sent out to the transmitter. Otherwise, the transmitter will receive a negative acknowledgment (NACK) signal which indicates that the received packet has been detected in error. Note that the transmitter is idle while waiting for the acknowledgment, and buffering is not necessary. The throughput efficiency of a SW-ARQ system can be derived as follows. First let us define the following parameters:

- T_{idle} , idle time which is the sum of the forward and reverse propagation delays and the time required for the receiver to determine if there are any errors in the received packet (processing delay);
- r , the data transmission rate in bits/second;
- n , number of bits in one data packet;
- P_r , probability of retransmission.

During one round-trip delay time (actual transmission plus idle time), the transmitter could have transmitted a total of $n + rT_{idle}$ bits if it did not stay idle. The average number of times a packet must be transmitted before it is accepted (positive acknowledgment) is:

$$N_{SW} = \sum_{i=1}^{\infty} i P_r^{i-1} (1 - P_r) = \frac{1}{1 - P_r} \quad (3.2)$$

While actually only n bits were accepted, the transmitter could have transmitted a total of $N_{SW}(n + rT_{idle})$ bits. Thus, the throughput efficiency is:

$$\eta = \frac{n}{N_{SW}(n + rT_{idle})} = \frac{1 - P_r}{1 + rT_{idle}/n} \quad (3.3)$$

Transmitted packets

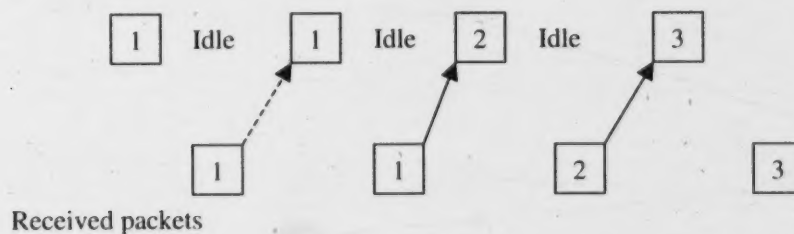


Fig. 3.5. SW-ARQ.

The SW-ARQ scheme is simple to implement, but its throughput is inherently low because of the idle time spent waiting for an acknowledgment for each transmitted packet. The processing delay is the dominant factor in the wireless cellular environment, which normally includes the processing time at both interleaver and deinterleaver, and the time for decoding, and it could be up to 100 ms for a typical CDMA channel [12]. For example, in the IS-95 system [3], there are only 184 information bits per data packet, and the basic data rate is 9600 bits per second, and idle time could be around 100 ms; that

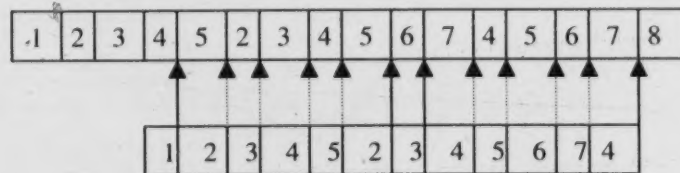
is $rT_{idle}/n \approx 5.2$. Thus even if the retransmission rate P_r is close to zero, η_{sw} is still very low. Hence the SW-ARQ has little attractive value for CDMA data transmission systems.

In the GBN-ARQ system, data packets are transmitted continuously. As shown in Fig. 3.6, the transmitter sends packets continuously until it receives a NACK, and then it will resend the requested packet and all subsequent packets (N packets in total). Clearly buffer must be provided at the transmitting end for these packets, but not necessarily at the receiving end. Note that N is the integer number of packets that contain a total of rT_{idle} information bits; that is,

$$N = \left\lceil \frac{rT_{idle}}{n} \right\rceil$$

where $\lceil x \rceil$ denotes the smallest integer larger than x .

Transmitted packets



Received packets

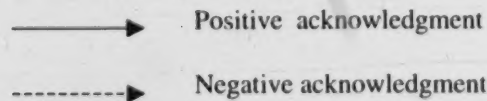


Fig. 3.6. GBN-ARQ with $N=4$.

In a GBN-ARQ system, for a packet to be successfully accepted, the average number of transmissions required is:

$$N_{GBN} = \sum_{i=0}^{\infty} (iN + 1)P_r^i(1 - P_r) = 1 + \frac{NP_r}{1 - P_r} \quad (3.4)$$

where the term $iN + 1$ is the number of transmitted packets corresponding to i negative acknowledgments. Obviously, the transmitter potentially could have transmitted a total of nN_{GBN} information bits during the time required for n information bits to be accepted.

Therefore, the throughput efficiency of a GBN-ARQ system is:

$$\eta_{GBN} = \frac{n}{nN_{GBN}} = \frac{1 - P_r}{1 + P_r(N - 1)} \quad (3.5)$$

Because of the continuous nature of transmission (no waiting time), the GBN-ARQ is more effective than the SW-ARQ, especially when the channel error rate is low (the effect of $P_r(N - 1)$ is insignificant). However, when the round-trip delay is large and the data transmission rate is high, which means that N is very large, the GBN-SRQ scheme becomes inefficient.

The most efficient and also complex scheme among the three basic ARQ schemes is the SR-ARQ. In the SR-ARQ system, data packets are also transmitted continuously. However differently from the GBN-ARQ, the transmitter only resends the packets that were detected with errors (negatively acknowledged) as shown in Fig. 3.7. The average number of transmissions before a packet is accepted, denoted by N_{SR} is the same as N_{SW} in (3.2). Thus, the throughput efficiency of an SR-ARQ system is simply:

$$\eta_{SR} = \frac{1}{N_{SR}} = 1 - P_r \quad (3.6)$$

The price for the high efficiency of the SR-ARQ scheme is the need of sufficient buffers at both the transmitting and receiving ends. We note that for the GBN-ARQ scheme, when $N = 1$, GBN-ARQ becomes equivalent to SR-ARQ. For example in (3.4), if $N = 1$,

$$N_{GBN} = \frac{1}{1 - P_r} = N_{SR}, \text{ and } \eta_{GBN} = \eta_{SR}.$$

Transmitted packets



Received packets

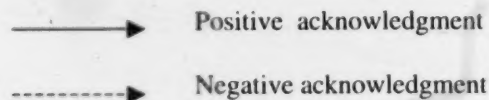


Fig. 3.7. SR-ARQ.

3.2.2.2 Hybrid ARQ

As briefly introduced in Section 3.1, a combination of the ARQ and FEC schemes, referred to as the hybrid ARQ [13]- [17], can provide very high reliability, and also high system throughput for data transmission. The function of the FEC is to correct packet errors, and thus reduce the retransmission rate. This naturally increases the system throughput. Retransmission schemes, on the other hand, are to make sure packets detected in error are retransmitted. This increases the reliability. Hence, a hybrid ARQ system provides higher reliability than a pure FEC, and a higher throughput than a pure

ARQ. There are two types of hybrid ARQ schemes. One is called the type I hybrid ARQ, and the other is the type II hybrid ARQ.

The type I hybrid ARQ uses either one code or concatenated codes (inner and outer codes). If one code is used, it is designed for simultaneous error correction and error detection. For the system using concatenated codes, the inner code is normally designed for error correction, and the outer code for error detection (of course, the outer code can also be designed for simultaneous error correction and error detection). The type I ARQ using one code, normally an (n, k) linear block code [18], has not been used in wireless cellular systems. The reason is that no efficient soft decision decoding algorithms are available, especially for large block linear codes. For example, in all three popular second generation cellular standards, IS-95, IS-136, and GSM, convolutional codes are used for channel coding. Even for the upcoming third generation systems, linear block codes are not used as inner codes of hybrid ARQ schemes. Although convolutional codes can also be used for both error correction and detection [19]-[20], these techniques are either not technically mature enough, or slow in decoding. A concatenated code performs the error correction and detection tasks separately. Hence, some effective soft decision techniques, such as the Viterbi decoding algorithm, can be used for error correction, and the syndrome computation using shift register circuits, can be used for error detection. Type I hybrid ARQ with a concatenated code is the focus of this chapter.

All type II hybrid ARQ schemes involve some form of packet combining. As opposed to the type I ARQ, the information from the received packet detected in error is not simply

discarded; it is later combined with received information from retransmitted packets, creating more reliable information which can be used for decoding. The interest in type II hybrid ARQ has led to a large number of publications [16], [21], [22], [23]. Nevertheless, the type II hybrid ARQ scheme is not an attractive scheme in commercial integrated voice/data cellular systems. The main reason is that the decoder chip for data users using the type II hybrid ARQ scheme is different from that for voice users, and hence data users can not share the same hardware with voice users. This increases the complexity of the design of the whole integrated system. Data users using the type I hybrid ARQ scheme may share the same decoding hardware with voice users in the integrated voice/data system, and this makes the type I hybrid ARQ scheme more attractive. Therefore, the type II hybrid ARQ system is not considered in this thesis.

3.2.2.3 Retransmission schemes

The retransmission scheme proposed here is negative acknowledgment (NACK) based which is in general terms similar to the scheme of the retransmission protocols of IS-99 [24]. That is, if the received frame is detected error free, no acknowledgment will be sent to the transmitter; only retransmission of the data frames in which errors were detected is requested. The basic protocols discussed above assume the feedback channel is error free. Actually since the logical feedback channel also uses a physical radio channel, it is certainly not error free. To handle the possibility that the retransmission request may never reach the transmitter, a retransmission timer is employed. That is, each time the receiver sends a retransmission request, a timer for that request is started. If a requested data frame is not received after its corresponding retransmission timer elapses, the

receiver shall send additional retransmission requests until it receives the packet. All packets have their own sequence numbers, so that they can be tracked by the transmitter and receiver. The additional details of retransmissions can be summarized as follows:

- The retransmission request frame is treated the same as the data frame on the downlink. It is error-detection and error-correction encoded, and if it is detected in error, the NACK is erased, and the transmitter continues to send new data frames.
- The complete receiver state information should be contained in each NACK message as proposed in [25]. That is, it should contain not only the sequence number of the data packet that has just been found in error, but also sequence numbers of all data packets that were detected in error and not received because of the feedback channel errors (NACK erased).
- If a data packet has already been accepted as error-free by the receiver, and the same packet arrives again because of incorrectly decoded NACK, this data packet shall be discarded.
- The transmitter should send the packets requested for retransmission with the highest priority.

The system implementation will also have to contain a time-out limit beyond which the packet is declared undeliverable (the number of erased NACKs received exceeds threshold). To illustrate how the NACK based protocol works, an example is depicted in Fig. 3.8, where round trip and processing delays are assumed to be four packets long. A brief description of this protocol's operation follows from Fig. 3.8. The mobile transmitter continuously sends new packets over the channel before receiving an error free NACK

packet, after which it resends the previously erroneously received packet #3, which is accepted by the receiver after the round trip and processing delay. Packet #4 is also detected in error, but since the transmitter fails to retransmit it because of the erroneous NACK packet, the NACK message for packet #7 requests retransmission of both packets #4, and #7. Note that in Fig. 3.8 packets 3, 4 and 7 were initially detected in error by the receiver.

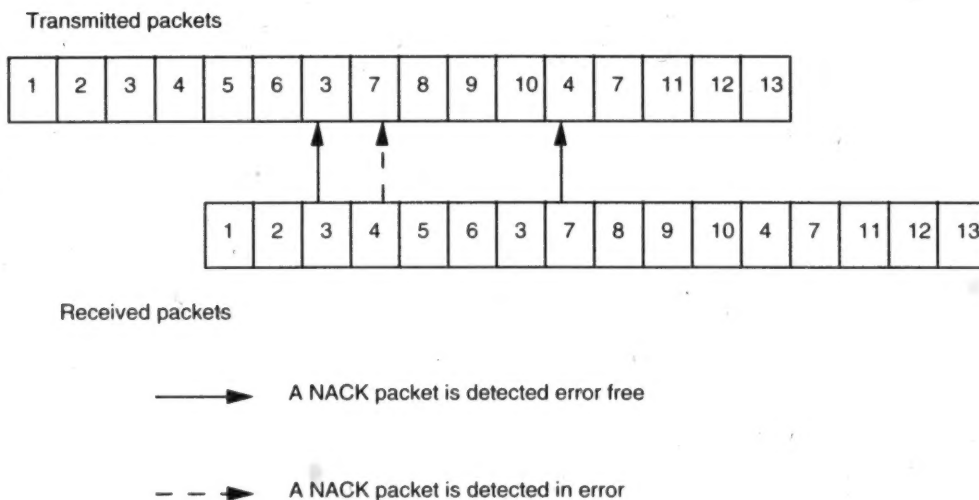


Fig. 3.8. Example of the operation of the NACK protocol.

3.3 Performance analysis

3.3.1 Error detection

The performance of an error detection code is measured by the probability of an undetected error at a given input bit error rate. As already mentioned in Section 3.1, internationally standardized CRC codes may not be suitable for very high-reliability data transmission in wireless systems, because with only up to 16 parity check bits they are

not able to ensure a sufficiently low probability of an undetected error at link bit error rates typical for mobile radio channels (e.g. $p = 10^{-3}$). The standard CRC codes are designed for links with initial bit error rates below 10^{-4} [4], whereas in uncoded wireless links normally $p = 10^{-1}$, and when very powerful FEC codes are used, typically $p = 10^{-3}$.

Consequently an (n, k) linear binary BCH code has been chosen for error detection. The selection of this code has been motivated by the required QoS (quality of service). Specifically, the code's probability of an undetected error has to be less than the accepted packet error rate tolerable by the data transmission system, otherwise, the percentage of the packets accepted by the receiver that contain one or more bit errors can be higher than the required minimum. In this subsection, we shall first introduce the proposed BCH error detection code. Then its fundamental performance characteristic, the undetected error probability, will be determined as a function of the channel bit error rate. In later sections it will be shown that the proposed code can guarantee on accepted packet error rate of less than 10^{-14} if the channel bit error rate is around 10^{-3} .

Before deriving the formula to compute the undetected error rate of an (n, k) linear block code, let us introduce the so-called MacWilliams identity [26]. Let $\{A_0, A_1, \dots, A_n\}$ where A_i is the number of code words of weight i be the weight distribution of an (n, k) linear block code, and $\{B_0, B_1, \dots, B_n\}$ be the weight distribution of its dual code. These weight distributions can be represented in polynomial form as follows:

$$A(z) = A_0 + A_1 z + \dots + A_n z^n \quad (3.7a)$$

$$B(z) = B_0 + B_1 z + \dots + B_n z^n \quad (3.7b)$$

Then $A(z)$ and $B(z)$ are related by the following MacWilliams identity:

$$A(z) = 2^{-(n-k)} (1+z)^n B\left(\frac{1-z}{1+z}\right) \quad (3.8)$$

An undetected error occurs only when the error pattern is identical to a nonzero code vector, hence, the probability of an undetected error for an (n, k) linear code used on a binary symmetric channel (BSC) is:

$$P_{ue} = \sum_{i=1}^n A_i p^i (1-p)^{n-i} \quad (3.9)$$

where p is the channel bit error rate (or channel transition probability). (3.9) can be further written in the following form:

$$P_{ue} = (1-p)^n \sum_{i=1}^n A_i \left(\frac{p}{1-p}\right)^i \quad (3.10)$$

Substituting $z = p/(1-p)$ in $A(z)$ of (3.7a), we have:

$$A\left(\frac{p}{1-p}\right) - A_0 = \sum_{i=1}^n A_i \left(\frac{p}{1-p}\right)^i \quad (3.11)$$

Note that the number of zero code words is 1; that is, $A_0 = 1$. Using the MacWilliams identity of (3.8), the right side of (3.11) is:

$$A\left(\frac{p}{1-p}\right) - 1 = 2^{-(n-k)} \frac{1}{(1-p)^n} B\left(\frac{1-p/(1-p)}{1+p/(1-p)}\right) - 1 = 2^{-(n-k)} \frac{1}{(1-p)^n} \sum_{i=0}^n B_i (1-2p)^i - 1 \quad (3.12)$$

The probability of an undetected error of (3.10) can thus be represented as:

$$P_{ur} = (1-p)^n \left(A \left(\frac{p}{1-p} \right) - A_0 \right) = 2^{-(n-k)} \sum_{i=0}^n B_i (1-2p)^i - (1-p)^n \quad (3.13)$$

The code proposed for further consideration in this work is the BCH(255, 231) primitive binary code. The code generator polynomial in octal form is 156720665 [1], [2]; that is, the generator polynomial has the following form:

$$156720665 \Rightarrow 001,101,110,111,010,000,110,110,101$$

$$g(x) = x^{24} + x^{23} + x^{21} + x^{20} + x^{19} + x^{17} + x^{16} + x^{15} + x^{13} + x^8 + x^7 + x^5 + x^4 + x^2 + 1 \quad (3.14)$$

The encoding process for any (n, k) systematic linear cyclic block code is implemented as follows [7]:

1. Multiply the information polynomial $M(x)$ by x^{n-k} .
2. Divide $x^{n-k}M(x)$ by $g(x)$ to obtain the remainder $r(x)$.
3. Add $r(x)$ to $x^{n-k}M(x)$.

For error detection, syndrome decoding [2] can be used here. To be specific, let us represent the received code vector $\mathbf{Y}$ by the polynomial $Y(x)$, and divide $Y(x)$ by the generator polynomial $g(x)$. If the remainder is zero, the code vector is declared error free; otherwise, an error for this code vector is declared.

The reason for choosing this code for our application is that on one hand a desired (184, 160) code can be obtained from it by shortening, and on the other hand, the shortened (184, 160) code has at least the same error-detecting capability as the original (255, 231) BCH code [1]. A computer search has been conducted to find the weight distribution of

the dual code of the BCH(255, 231) primitive binary code. The results of the search are presented in Table 3.1¹, and P_{ue} versus the channel error probability is given in Table 3.2.

Table 3.1. Weight distribution of the dual code of the BCH (255, 231) binary code

i	B_i
0	1
96	10710
112	1370880
120	3588224
128	7568655
136	3166080
144	1066240
160	6426

It is known that the probability of an undetected error for a double-error correcting primitive BCH code of length $2^m - 1$ is upper-bounded by 2^{-2^m} for $p \leq 0.5$ [1]. For triple-error-correcting primitive BCH codes of length $2^m - 1$, such as the (255, 231) code ($m = 8$), it has been demonstrated in [27] that when the channel bit error rate p is within a small interval around 0.0994196, the undetected error probability is slightly larger than

¹ The computer search is optional, since the weight distribution for a double-error-correcting or triple-error-correcting primitive BCH code can also be computed from one of the tables given in [1].

the $2^{-24} (\approx 5.9604645 \times 10^{-8})$. Hence, the conjecture [1] that the probability of an undetected error for any primitive BCH code is upper bounded by $2^{-(n-k)}$ is not exactly correct. Our results obtained from (3.13) also confirm this behavior. However, the application of (3.13) to the (255, 231) code ($n-k=24$) reveals that P_{ue} is upper bounded by $(2^{-24} + 7.5443 \times 10^{-15}) \approx 5.9604652 \times 10^{-8}$, which is only marginally (by about $10^{-5}\%$) larger than 2^{-24} .

Table 3.2. Probability of undetected error (P_{ue})

Channel bit error rate (p)	Undetected error rate (P_{ue})
0.5	5.9604645E-08
0.2	5.9604645E-08
0.1	5.9604645E-08
0.09942	5.9604652E-08
0.02	1.62476285E-08
0.01	1.00146638E-09
0.009	5.92001364E-10
0.007	1.5603576E-10
0.005	2.2716884E-11
0.003	9.7799546E-13
0.001	8.8751611E-16
0.0008	4.1102331E-16

3.3.2 Error correction

The convolutional encoder is a finite-state machine (FSM). It contains memory elements whose contents and the inputs at that time determine the encoder outputs. The state of the convolutional encoder is defined as its memory contents. In the system considered in this chapter (see Fig. 3.1), error correction capability is provided by convolutional coding with soft-decision Viterbi decoding. To be specific, a rate 1/3 convolutional code with a constraint length 9 is used. The three connection vectors are: $\mathbf{g}_0 = 557$ (octal), $\mathbf{g}_1 = 663$ (octal), and $\mathbf{g}_2 = 711$ (octal). The detailed structure of this convolutional encoder is shown in Fig. 3.9. The encoder consists of an $m = 8$ stage shift register and 3 modulo-2 adders.

The Viterbi algorithm [28]-[30] is a maximum likelihood decoding algorithm for convolutional codes [30], which implies that the decoder output sequence is always the one associated with the largest value of the log-likelihood function. To explain the operation of the Viterbi algorithm, it is convenient to describe the convolutional code with its trellis diagram. A trellis diagram of a convolutional encoder is the two-dimensional representation of the operation of the encoder; that is, it covers all possible state transitions in time. The complexity of a trellis diagram depends on the number of states of an encoder. For example the encoder in Fig. 3.9 has a total of $2^8 (m=8)$ or 256 states. It would be difficult to draw a trellis diagram for this encoder. We will instead consider a simple convolutional encoder and draw its trellis diagram to explain the Viterbi algorithm. The encoder is given in Fig. 3.10. This rate 1/3 convolutional encoder with $m = 2$ has four states: $S_0 = 00$, $S_1 = 10$, $S_2 = 01$, and $S_3 = 11$. Its trellis diagram is

shown in Fig. 3.11. The label '1/111' next to the trellis branch representing transition from state S_0 to state S_1 indicates input 1 and the corresponding output 111. Any specific encoded sequence is represented by a unique path through the trellis.

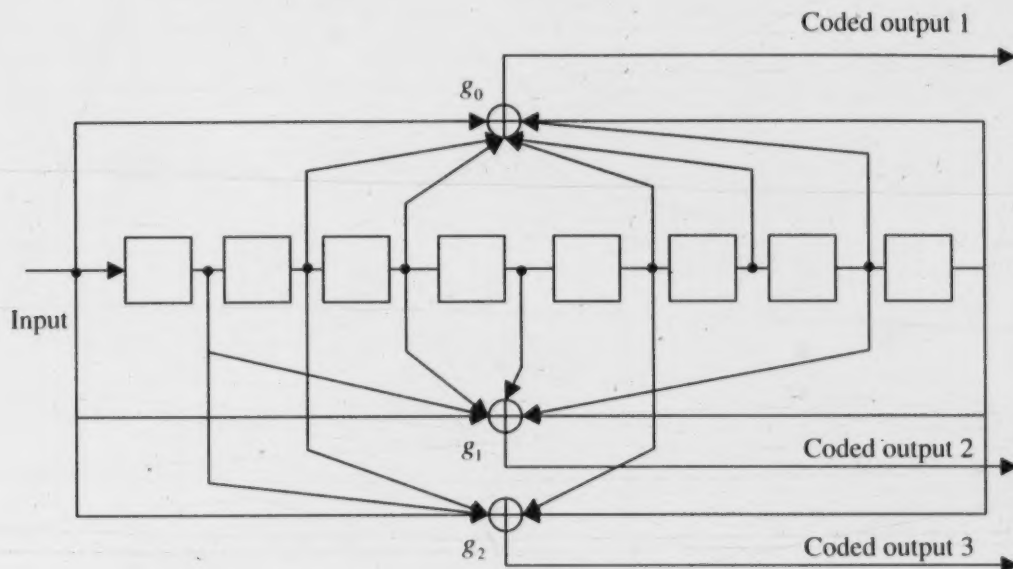


Fig. 3.9. Constraint length 9, rate 1/3 convolutional encoder.

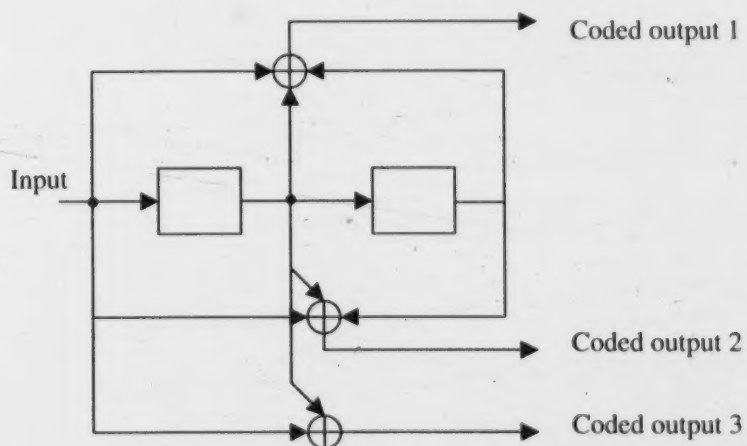


Fig. 3.10. A constraint length of 3, rate 1/3 convolutional encoder.

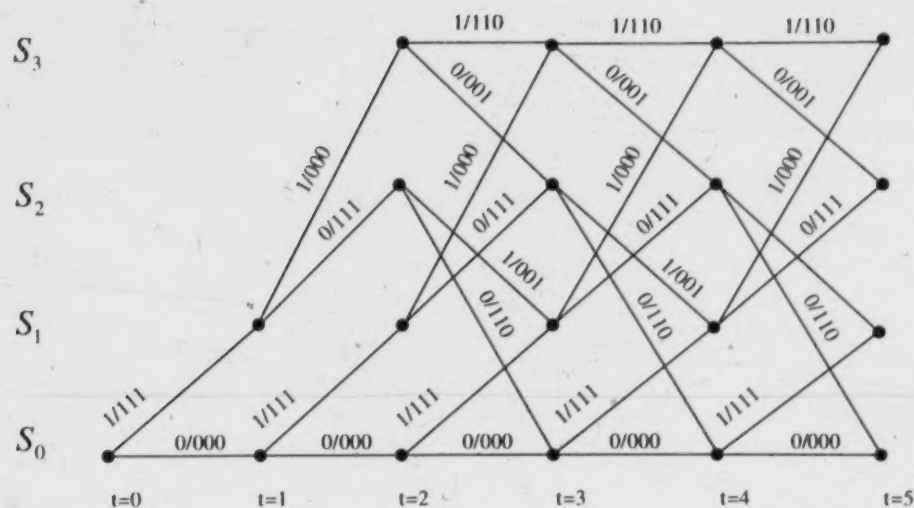


Fig. 3.11. Trellis diagram for the encoder in Fig. 3.10.

Now we can explain the Viterbi algorithm by examining Fig. 3.12. We assume that an information sequence $\mathbf{u} = (u_0, u_1, \dots, u_{N-1})$ is encoded by a convolutional encoder with coding rate $1/n$ in Fig. 3.12 into a coded sequence $\mathbf{v} = (v_0^1, v_0^2, \dots, v_0^n, v_1^1, \dots, v_{N+m-1}^n)$. The decoder produces an estimate $\hat{\mathbf{u}} = (\hat{u}_0, \hat{u}_1, \dots, \hat{u}_{N-1})$ based on the received sequence $\mathbf{r} = (r_0^1, r_0^2, \dots, r_0^n, r_1^1, \dots, r_{N+m-1}^n)$. The Viterbi algorithm chooses $\hat{\mathbf{u}}$ as the information sequence $\mathbf{u}$ to maximize the log-likelihood function $\log P(\mathbf{r} | \mathbf{v})$. Assuming the channel is memoryless, $\log P(\mathbf{r} | \mathbf{v})$ has the following form:

$$\log P(\mathbf{r} | \mathbf{v}) = \sum_{i=0}^{N+m-1} \left(\sum_{j=1}^n \log P(r_i^j | v_i^j) \right) \quad (3.15)$$

$\log P(\mathbf{r}|\mathbf{v})$ is called the path metric, terms $\log P(r_i^j | v_i^j)$ in (3.15) are bit metrics, and terms $\sum_{j=1}^n \log P(r_i^j | v_i^j)$ are branch metrics. The l -th partial path metric for a path is defined as the sum of the branch metrics for the first l branches that a path traverses, which can be expressed as $\sum_{i=0}^{l-1} \left(\sum_{j=1}^n \log P(r_i^j | v_i^j) \right)$.

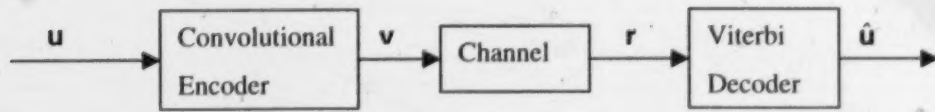


Fig. 3.12. Block diagram of a communication system using a convolutional code.

The metric information for the Viterbi decoder considered here is the square of the envelope detector output [11], as shown in Fig. 3.4. All M combined outputs in Fig. 3.3 are used for decoding. We will justify in the following as in [11] that the square-metric is indeed the optimal maximum likelihood (ML) metric for multipath Rayleigh fading channels with diversity combining.

First, we normalize the received squared output R in Fig. 3.4 by N_0 , the background noise power spectral density; that is $r = R / N_0$. Note that R in Fig. 3.4 is the sum of squares of two Gaussian random variables. For an AWGN channel, the decision variable when signals are present is noncentral chi-square-distributed with two degrees of

freedom, and when signals are absent it is central chi-square-distributed. For the flat Rayleigh fading channel, the respective probability density functions are given by [35]:

$$p_1(r) = e^{-r/\epsilon_L + 1}; \quad \text{Signal present} \quad (3.16)$$

$$p_0(r) = e^{-r}; \quad \text{Signal absent} \quad (3.17)$$

where ϵ_L is the average SIR per branch. For the receiver with L -branch diversity, the received signal is the sum of the corresponding squared magnitudes of the L outputs. The probability density functions of the sum of the L squared outputs in the multipath Rayleigh fading channel are given by [11], [35]:

$$p_1^L(r) = \frac{r^{L-1} e^{-r/\epsilon_L + 1}}{(L-1)!(\epsilon_L + 1)^L} \quad (3.18)$$

$$p_0^L(r) = \frac{r^{L-1} e^{-r}}{(L-1)!} \quad (3.19)$$

From (3.18) and (3.19), the optimal ML metric is simply:

$$ML(r) = \ln \frac{p_1^L(r)}{p_0^L(r)} = r \left(1 - \frac{1}{\epsilon_L + 1}\right) - L \ln(\epsilon_L + 1) \quad (3.20)$$

Since $-L \ln(\epsilon_L + 1)$ is a common term and $1 - 1/(\epsilon_L + 1)$ is a common factor for all transitions, the optimal metric is simply r , the sum of the corresponding squared magnitudes of the L envelope detector outputs.

The Viterbi algorithm involves a search through the trellis for a sequence with the largest metric. When the convolutional code has a frame structure (frame length equal to $N+m$,

including m flushing zeros) the VA decoding can be described as follows [1], [2] (symbols used are the same as in Fig. 3.11):

1. Beginning at time $t = 1$, compute the partial metric for every path entering each state. Store the path with the largest metric (the survivor) and its metric for each state.
2. Increase t by 1. Compute the partial path metric for all the paths entering a state by adding the branch metric entering that state to the metric of the connecting survivor at the preceding time unit. For each state, store the path with the largest metric (the survivor), together with its metric, and delete other paths. Note that ties can be broken by the flip of a coin.
3. If $t < N + m$, repeat step 2. Otherwise, stop.
4. Start at state S_0 , time $t = N + m$ and follow the surviving branches backward through the trellis to state S_0 at time $t = 0$, a unique path can be found which corresponds to the maximum likelihood sequence block.

In general, a frame error rate (FER) is not a suitable performance measure for convolutional codes, since unlike a block code, a convolutional code in general does not have a frame structure. However, in wireless systems, information sequences are grouped into frames before encoding. At the end of each frame the convolutional encoder is reset to zero by flushing zeros appended at the end of each information frame. Thus in this case, it is quite natural to express the convolutional code's performance by its frame error rate (FER). An exact closed form formula for the FER as a function of bit energy-to-noise ratio, code constraint length, etc. is not available, and usually it is obtained by simulation as done in this work.

3.3.3 Power control

In CDMA systems, all mobile stations share the same radio frequency channel. The base station may receive stronger signals from mobile stations close to the base station than those far away from the base station (near-far problem). Also, mobile stations in deep fades need more transmit power, since if a mobile station's signal arrives at the base station with too low a received power value, the bit error rate is too high to permit high quality communications. Hence it is necessary to control the transmitted power of each mobile station to eliminate near-far problem and SIR fluctuation due to shadowing and small scale fading and keep SIR as low as possible (for a given QoS) to reduce interference. By doing that, the CDMA system capacity can be maximized.

Power control is a technique used to overcome the near-far problem, and mitigate small scale fading. By using a power control mechanism, regardless of a mobile station's position or propagation loss, each mobile station's signal will be received at the cell at the same level. Power control in the CDMA system is further divided into two categories. One is called open-loop power control, which attempts to have all mobile stations transmitted signals arrive at the base station with the same nominal power level. Open loop power control is mainly used to compensate for large-scale path loss (coarse power adjustment), while close-loop power control is used to provide rapid corrections to the mobile station's open loop estimate to maintain the optimum transmit power (fine power adjustment). It is used to reduce small-scale fading.

The large-scale path loss $PL(d)$ where d is the transmitter-receiver separation distance, is proportional to [31]:

$$PL(d) \propto 10n \log(d) \quad (3.21)$$

where n is the path loss exponent which indicates the rate at which the path loss increases with distance (The typical value of n in the urban area cellular radio channel is around 2.7~3.5 [31]). Open loop power control is achieved by measuring the power level of the pilot signal transmitted by the base station; the stronger the received pilot, the lower the mobile station's transmitted power. Closed-loop power control is achieved through the measurement of the SIR received from a given mobile at the base station. Closed-loop power control is more difficult than open-loop power control. The algorithm described below is for closed-loop power control, which includes both inner and outer loops. Closed-loop power control bit is generated every T_p by the base station using the M Walsh coefficients of Fig. 3.3. The received signal and noise are estimated by the inner loop [32]:

$$\bar{E}_s = \frac{1}{n_s} \sum_{j=1}^{n_s} \max_k \{R_{j,k}\} - \frac{\bar{n}^2}{M-1} \quad (3.22)$$

$$\bar{n}^2 = \frac{1}{n_s} \sum_{j=1}^{n_s} \sum_{k \neq k_{\max}}^M R_{j,k} \quad (3.23)$$

where $k_{\max}$ is the index of the largest decision variable for the given symbol interval, n_s is the number of symbols in the power control measurement interval of T_p ms, and $R_{j,k}$ is the k -th decision variable at the output of the diversity combiner in Fig. 3.3 during the j -th symbol interval. The estimated SIR in (3.22) and (3.23) is compared with a threshold which is set by the outer loop to maintain a desired received frame error rate, and

accordingly a power control bit is generated to increase or reduce the transmitted power by a fixed quantity.

3.3.4 Throughput and reliability

We begin the analysis by first assuming the feedback channel is noise free which means all negative acknowledgment messages can be correctly received by the transmitter. We shall discuss the effect of feedback channel errors later.

The BCH code and the convolutional code can be viewed as the outer and inner codes respectively in the concatenated coding system. Let us define the following probabilities:

- P_r is the probability of a retransmission request, or that the error in the received frame is detected;
- P_c is the probability that the received frame contains no error;
- P_{ue} is the probability of undetected error for the BCH code;
- $P(E)$ is the probability that a packet containing errors is accepted by the receiver.

Clearly $P_c = 1 - P_F$. Also, since $P_c + P_{ue} + P_r = 1$, P_r can be expressed as a function of P_F :

$$P_r = P_F - P_{ue} \quad (3.24)$$

Therefore the accepted packet error rate can be readily obtained as:

$$P(E) = \sum_{i=0}^{\infty} P_{ue} P_r^i = \frac{P_{ue}}{1 - P_r} = \frac{P_{ue}}{1 - P_F + P_{ue}} \quad (3.25)$$

From (3.25), both lower and upper bounds on $P(E)$ can be written as:

$$P_{ue} < P(E) < \frac{P_{ue}}{1 - P_F} \quad (3.26)$$

As shown in [5], the undetected error probability P_{ue} of the standard CRC codes is around 10^{-9} when the channel bit error rate is 10^{-3} , which usually corresponds to the frame error rate of 10^{-2} for short packets of about 100~200 bits. If these codes are used for error detection, from (3.26) it is clearly seen that the accepted packet error rate $P(E)$ will be higher than 10^{-9} . On the other hand, when the channel error rate is 10^{-3} , P_{ue} of the (184,160) code is less than 10^{-15} (see Table 3.2), and the accepted packet error rate $P(E)$ is about 10^{-15} . Hence, the standard CRC codes can only be used in case that the accepted packet error rate around $10^{-8} \sim 10^{-9}$ is acceptable. Inequality (3.26) should be used as the criterion for designing error detection codes for packet data transmission.

The throughput efficiency of the system in Fig. 3.1, defined as the ratio of the average number of frames successfully accepted by the receiver to the total number of the frames that have been transmitted, can be computed by first calculating the average number of times (N_r) that the frame must be transmitted before it is accepted:

$$N_r = \sum_{i=1}^{\infty} i(1 - P_r)P_r^{i-1} = \frac{1}{1 - P_r} = \frac{1}{1 - P_F + P_{ue}} \quad (3.27)$$

The throughput efficiency η is therefore:

$$\eta = \frac{1}{N_r} = 1 - P_F + P_{ue} \quad (3.28)$$

As mentioned earlier, in the above analysis a noise-free feedback channel has been assumed. Although feedback errors do occur, we will show that their effect may be ignored. As described in Section 3.2.2.3, each NACK message contains error detection

bits, and sequence numbers of all packets detected in error by the receiver. The probability of sending wrong packets by the transmitter is negligible, since the transmitter will disregard the NACK if it is detected in error. Let P_{FF} denote the packet error rate in the feedback channel. Typically with average 6 dB signal-to-interference ratio, P_{FF} around 10^{-2} can be achieved as demonstrated in [32]. In the ideal case, as long as the last NACK is successfully accepted by the transmitter, all erroneous packets will be retransmitted, and the throughput expression (3.28) will still hold. In practice, a NACK time-out timer can be implemented. If any data packet requested does not reach the receiving end before the timer expires, the packet is declared undeliverable at this layer. Let P_{ud} denote the probability of a packet being declared undeliverable. Let us assume forward and feedback CDMA channels having the same probability of the packet error ($P_F = P_{FF}$), all packets sent being independent of each other, and define:

- E_j , the event that the missing frame is not received correctly up to the j -th retransmission attempt.
- w , the maximum number of retransmission attempts.

Then we have [12]:

$$P(E_j) = \prod_{i=1}^j [1 - (1 - P_F)^2] \quad (3.29)$$

P_{ud} is thus given by:

$$P_{ud} = P_F \prod_{j=1}^w P(E_j) = P_F [1 - (1 - P_F)^2]^{w(w+1)/2} \quad (3.30)$$

P_{ud} versus P_F is plotted in Fig. 3.13. It is seen that in the normal case when the link frame error rate is around 10^{-2} , and $w = 3$ (default value in IS-99 [24]), the probability of a packet being declared undeliverable is about $10^{-12} \sim 10^{-13}$, which is negligible.

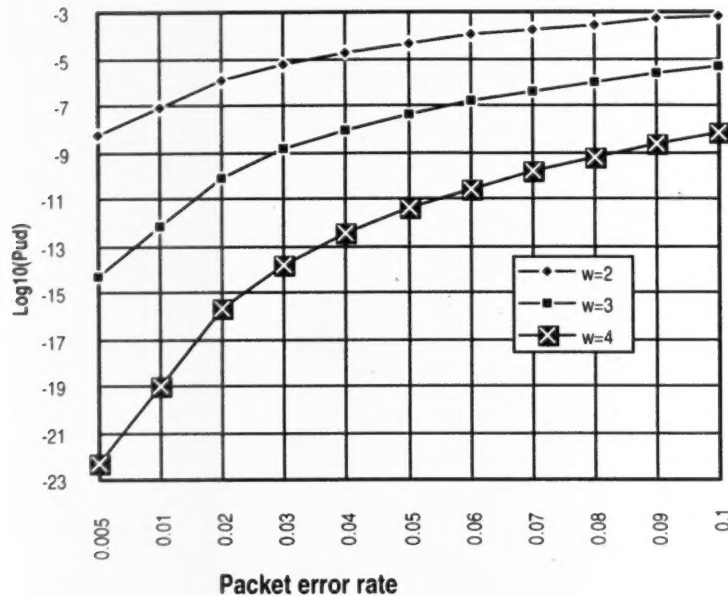


Fig. 3.13. Probability of a packet being declared undeliverable (P_{ud}) versus packet error rate (P_F).

3.4 Optimum signal-to-interference ratio assignment

It is quite clear that for packet data transmission, the more retransmissions, the more interference to the system is generated by data users, and minimizing the total interference (including the effect of retransmissions) leads to increased system capacity for an interference-limited CDMA system. In this subsection, we will show that

minimizing the cumulative SIR of a fully loaded system is equivalent to minimizing the total interference, and thus it increases the system capacity. Based on this argument, methods of finding the optimum SIR are introduced.

For a multi-service DS-CDMA wireless network, without considering interference from other cells, blocking as defined in [33] can be shown to occur when

$$\sum_{i=1}^{K_v} E_{b,vi} r_{vi} + \sum_{i=1}^{K_d} N_r(\epsilon_{di}) E_{b,di} r_{di} + N_0 W > I_0 W \quad (3.31)$$

where W is the spread-spectrum system bandwidth, N_0 is the background noise power spectral density, I_0 is the maximum total acceptable interference spectral density, K_v and K_d are numbers of voice and data service users per cell, $E_{b,vi}$, $E_{b,di}$ are bit energies of the i -th voice and data user with rates r_{vi} , r_{di} respectively, and $N_r(\epsilon_{di})$ is the average number of times that a data packet must be transmitted before it is accepted, which is a function of the SIR; $\epsilon_{di} = E_{b,di} / I_0$ is the SIR for the i -th data user. Note that in the above expression we have taken into account messages retransmitted due to errors, and the total (cumulative) interference generated by a data user is calculated as the actual link transmitted power times the average number of transmissions $N_r(\epsilon_{di})$.

Inequality (3.31) can be simplified to yield:

$$\sum_{i=1}^{K_v} \epsilon_{vi} r_{vi} + \sum_{i=1}^{K_d} N_r(\epsilon_{di}) \epsilon_{di} r_{di} > W(1 - \xi) \quad (3.32)$$

where $\xi = N_0/I_0$, and $\epsilon_{vi} = E_{b,vi}/I_0$ is the SIR for the i -th voice user; ϵ_{vi} , ϵ_{di} , K_v , and K_d are random variables.

From (3.32) it is clear that if we minimize each data subscriber's average cumulative SIR, $N_r(\bar{\epsilon}_{di})\bar{\epsilon}_{di} = \Omega(\bar{\epsilon}_{di})$, the network can accommodate more users than for an arbitrarily chosen $\Omega(\bar{\epsilon}_{di})$; $\bar{\epsilon}_{di}$ denotes the average value of ϵ_{di} . Obviously there must be an optimum average SIR per transmission, $\bar{\epsilon}_{di}$, which minimizes $\Omega(\bar{\epsilon}_{di})$ and leads to the increased system capacity, while reducing somewhat the data throughput per individual user. One method of finding the optimum average signal-to-interference ratio calls for differentiating $\Omega(\bar{\epsilon}_{di})$, and equating it to zero:

$$\frac{d}{d\bar{\epsilon}_{di}}(\Omega(\bar{\epsilon}_{di})) = N_r(\bar{\epsilon}_{di}) + \bar{\epsilon}_{di} \frac{d}{d\bar{\epsilon}_{di}}(N_r(\bar{\epsilon}_{di})) = 0 \quad (3.33)$$

Unfortunately, a closed form of the $N_r(\bar{\epsilon}_{di})$ can not be readily determined. A practical approach is to plot $\Omega(\bar{\epsilon}_{di})$ versus $\bar{\epsilon}_{di}$ using simulation results, and subsequently find the optimum average signal-to-interference ratio from the curve. Alternatively, we may approximate the retransmission request probability P_r as [34]:

$$P_r \approx a \exp(-b\bar{\epsilon}_{di}) \quad (3.34)$$

where a and b are constants, and their values depend on the modulation and coding schemes. Using (3.27), we have $N_r(\bar{\epsilon}_{di}) = \frac{1}{1-P_r} \approx \frac{1}{1-ae^{-b\bar{\epsilon}_{di}}}$. Hence, from (3.33), it can

be shown readily that the optimum SIR satisfies the following equation:

$$1 - ae^{-b\bar{E}_d} - ab\bar{E}_d e^{-b\bar{E}_d} \approx 0 \quad (3.35)$$

These two methods will be employed in the next section in order to obtain the optimum average SIR.

3.5 Simulation study

The data transmission system depicted in Fig. 3.1 has been extensively simulated in order to verify its performance, and to obtain the optimum received SIR per packet transmission. The performance measures are the accepted packet error rate and the throughput efficiency versus the equivalent despread received SIR. The effect of the Doppler shift on the system performance is also investigated. The accepted packet error rate, $P(E)$, and throughput efficiency, η , can be obtained using the equations (3.25), and (3.28), respectively. The throughput efficiency can also be obtained entirely by simulation. In equations (3.25) and (3.28), P_{ue} is available from (3.13), but P_F (the frame error rate of the inner, convolutionally encoded link in Fig. 3.1) can only be obtained from simulations, which are run at different received SIRs and Doppler shifts. Naturally, also the bit error rate, p , of the inner link, necessary in (3.13) to determine P_{ue} , needs to be obtained from simulations. Thus, there are two independent simulations. First is the simulation of the entire system in Fig. 3.1; the simulation results are used to determine the system throughput, and also the $\Omega(\bar{E}_d)$ curve necessary to determine the optimum SIR assignment. Second is the simulation of the inner, convolutionally encoded and 64-ary orthogonally modulated link in Fig. 3.1. It is necessary to find the bit error rate, p , and the frame error rate, P_F , at the output of the convolutional decoder versus the

required equivalent despread SIR. The values of p and P_F are then used in equations (3.13), (3.25) and (3.28) to determine $P(E)$ and η . Throughput efficiency calculated from (3.28) will be compared with the simulation results of the whole system in order to verify the analysis in Section 3.3.4.

In the first simulation, a total of 24000 frames have been simulated. Each frame contains 160 data bits, which are encoded into 184 code bits by the BCH encoder. 184 code bits plus 8 zero tail bits (a total of 192 bits) are further encoded into 576 code bits by the 1/3 convolutional encoder with memory size 8 (constraint length 9). The convolutionally encoded bits are grouped and mapped into 96 Walsh symbols. The interleaver size for Walsh symbol-by-symbol interleaving is 16×6 . We consider the system bandwidth of 1.25 MHz, and use a three path frequency selective channel model. The relative powers of the three paths are (0, -3, -6) dB. In the second independent simulation, we are only interested in the performance of the inner link, so error detection codes are not used. The following assumptions are made in the simulations:

- Perfect chip synchronization and carrier recovery;
- The fast power control processing and transmission delay equal to $2T_p$, where T_p is the duration of 6 Walsh symbols;
- The power control step size is set at 1 dB;
- The power control bit error rate of 10%;
- Each resolvable path experiences independent Rayleigh fading;
- Doppler frequencies of 15, and 5 Hz are simulated. They correspond to the mobile station's speeds of 18 and 6 km/h, respectively, at the 900 MHz carrier frequency.

Simulations are first performed to determine the overall system performance in terms of its throughput efficiency and the accepted packet error rate. Throughput efficiencies versus the equivalent despread received SIR are plotted in Fig. 3.14. The simulations have not produced any accepted packets with errors. This is understandable, since at the SIR values considered (SIR 3.5 dB), the accepted packet error rate, $P(E) < 1.1 \times 10^{-7}$, and only 24000 frames were simulated. It is apparent that increasing Doppler shift causes some reduction of the system throughput for SIRs in the range of 4.0–7.0 dB. It is also evident that at $f_D = 5$ Hz, the SIR required to achieve 99% throughput efficiency is around 6 dB, and at $f_D = 15$ Hz, the required SIR is around 7 dB. In data communications applications, it is probably reasonable to assume that the mobile station moves slower than in the case of voice users.

Simulations are then conducted to see the performance of the inner convolutional coding system with 64-ary orthogonal signaling. Simulation results for BER (p) and FER (P_F) versus the equivalent despread received SIR are shown in Figures 3.15 and 3.16. These results are then used to compute the accepted packet error rate in (3.25) and the throughput efficiency in (3.28). It is evident from the simulations that the required SIR to achieve 10^{-3} bit error rate and 10^{-2} frame error rate is around 6.3 dB when the Doppler shift is 15 Hz, and 5.1 dB when Doppler shift is 5 Hz.

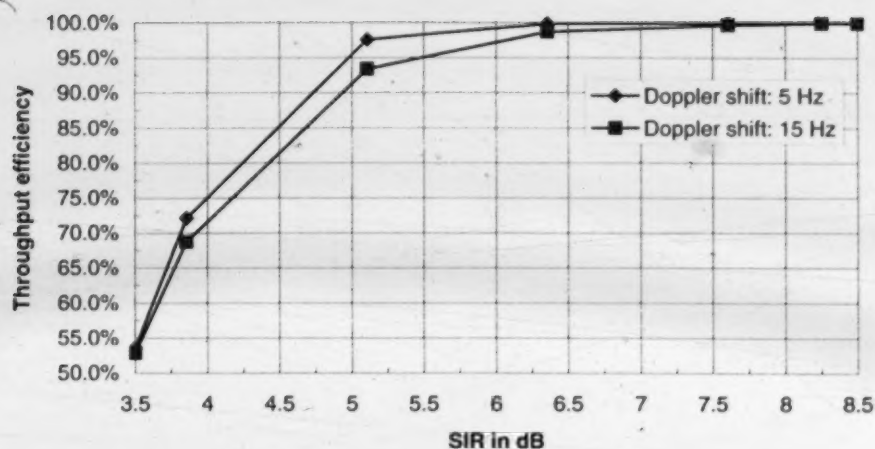


Fig. 3.14. Throughput efficiency versus SIR.

Using the simulation results plotted in Figures 3.15 and 3.16, we can compute the accepted packet error rate $P(E)$ from (3.25), and plot it versus the SIR as shown in Fig. 3.17; note that the BER vs. SIR in Fig. 3.15 is used as the channel bit error rate p to determine P_{uc} from (3.13). The results in Fig. 3.17 are consistent with our analysis in Section 3.3.3, where we claimed that the accepted packet error rate is around 10^{-15} for the channel bit error rate 10^{-3} , which corresponds to 5.1 dB for the Doppler shift 5 Hz, and 6.3 dB for the Doppler shift 15 Hz, respectively.

The results in Figures 3.15 and 3.16 are also used to compute the throughput efficiencies from (3.13) and (3.28), and compare them with the simulation results shown in Fig. 3.14. The comparison for the two selected Doppler shifts is presented in Figs. 3.18 and 3.19, where analysis refers to the results obtained from (3.13) and (3.28), and simulation results

are the same as those in Fig. 3.14. Clearly, the simulation results agree well with the analytical ones.

As stated in Section 3.4, the optimum average SIR per packet transmission can be found by plotting the average cumulative SIR versus the equivalent despread SIR per transmission, as done in Fig. 3.20. It is observed from the figure that the optimum SIR which minimizes the cumulative SIR is around 5.1 dB for both Doppler shifts of 5 and 15 Hz. The throughput corresponding to this optimum SIR is about 98% for Doppler shift of 5 Hz, and around 94% for Doppler shift of 15 Hz. As seen in Fig. 3.20, when the SIR per transmission exceeds 5.1 dB, the slope of the curve becomes steep, which implies that further increase in throughput efficiency will cause a significant reduction in the number of data users the system can accommodate.

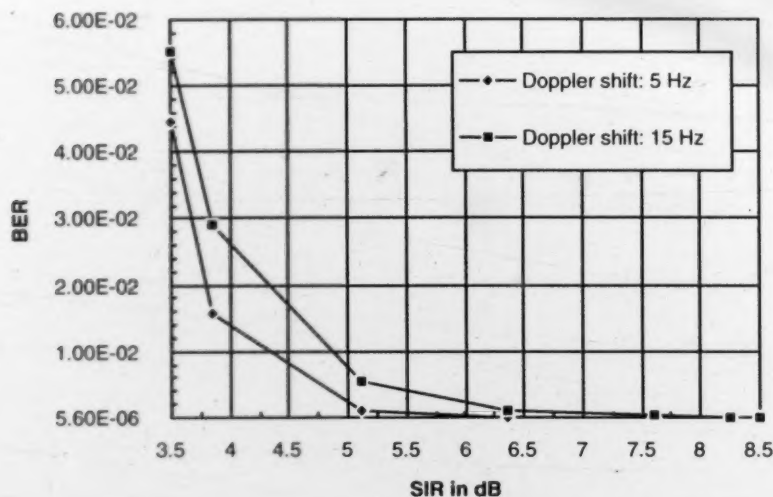


Fig. 3.15. BER (p) versus SIR for the 64-ary orthogonal signaling with rate 1/3 convolutional coding.

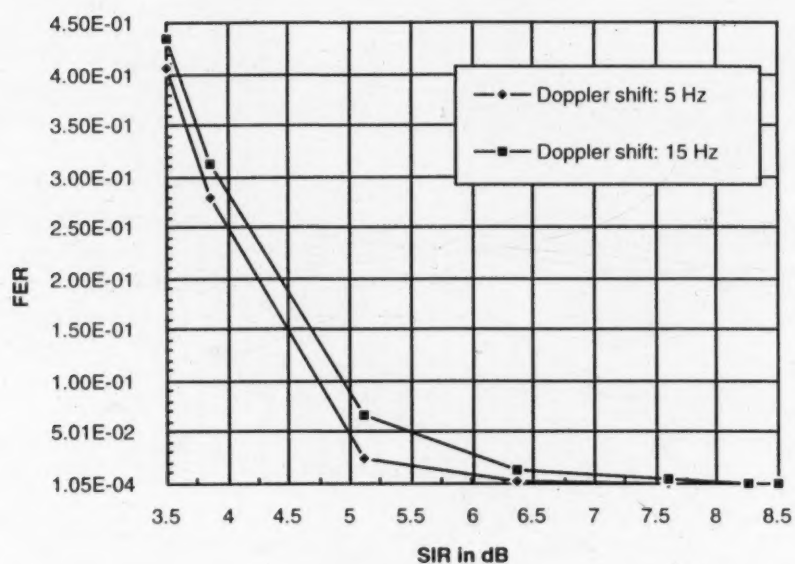


Fig. 3.16. FER (P_r) versus SIR for the 64-ary orthogonal signaling with rate 1/3 convolutional coding.

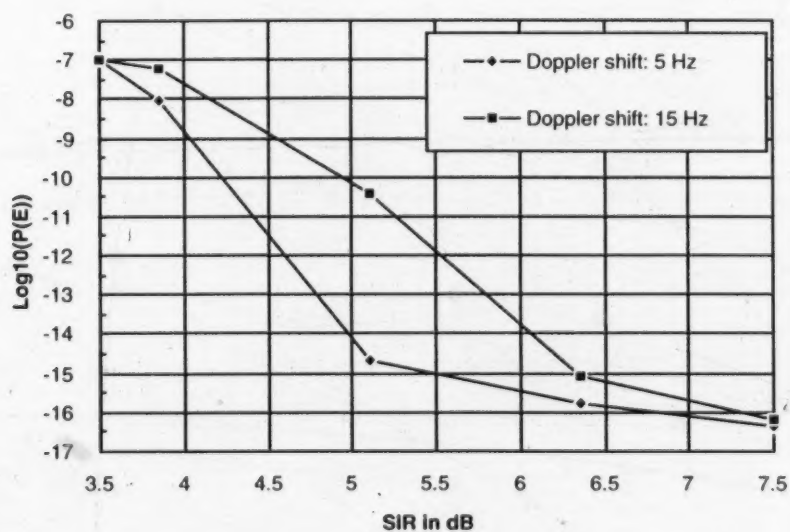


Fig. 3.17. Accepted packet error rate $P(E)$ versus SIR.

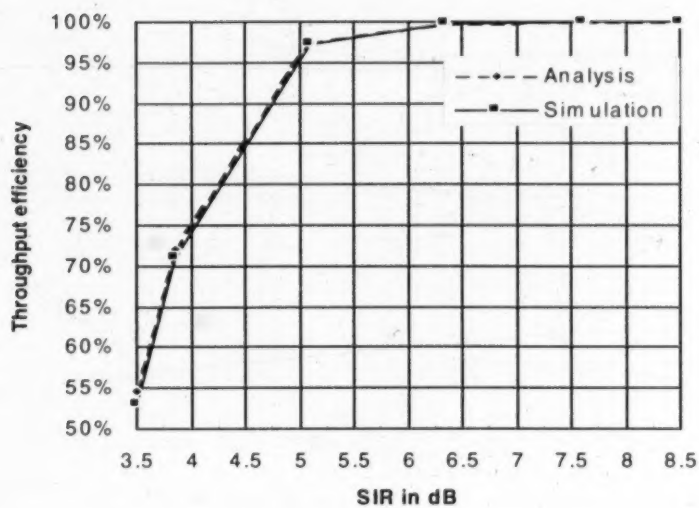


Fig. 3.18. Comparison of throughput efficiency vs. SIR for Doppler shift of 5 Hz.

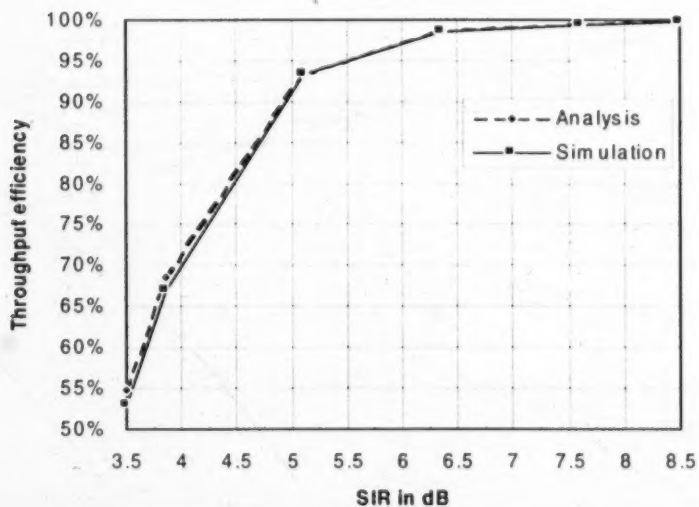


Fig. 3.19. Comparison of throughput efficiency vs. SIR for Doppler shift of 15 Hz.

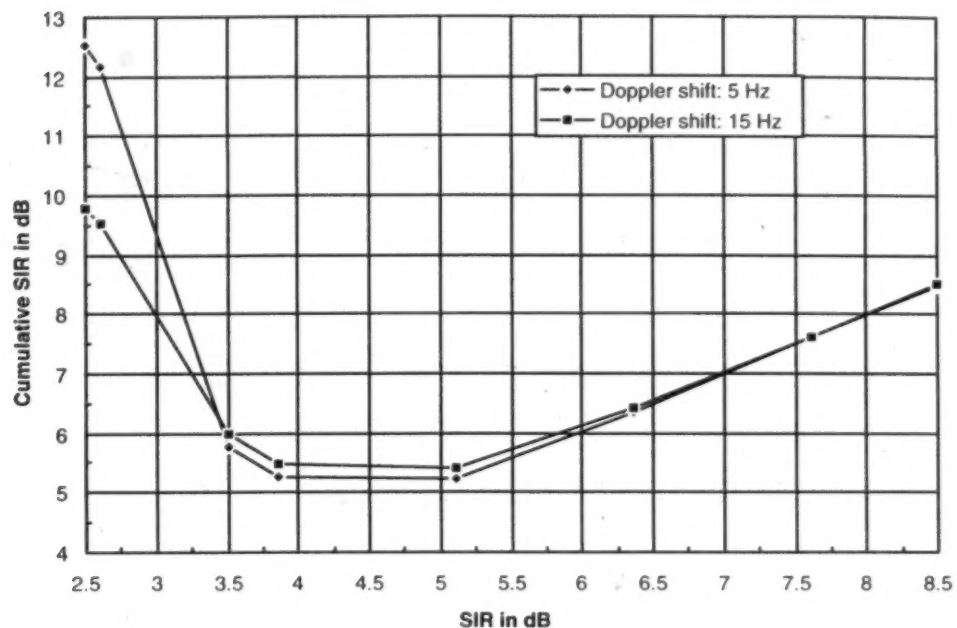


Fig. 3.20. Average cumulative signal-to-interference ratio versus SIR per transmission.

Another method of finding the optimum SIR per transmission relies on the approximation (3.35). The values of a and b are chosen so that the values of the probability of retransmission, $P_r = P_F - P_{ue}$, calculated from (3.34) are close to the ones obtained from simulations. Substitution of these values of a and b in (3.35) yields the approximately optimum SIRs for both Doppler shifts of 5 and 15 Hz, as listed in Table 3.3. These approximately optimum SIRs are close to those obtained from Fig. 3.20. From both Fig. 3.20 and Table 3.3, we also conclude that the optimum SIR per transmission is quite insensitive to the Doppler shift.

Table 3.3. Parameters and optimum SIRs

Doppler shift	5 Hz	15 Hz
<i>a</i>	54.75	8.323
<i>b</i>	2.1723	1.3498
Optimum SIR	4.45 dB	4.48 dB

3.6 Conclusions

Negative acknowledgment based hybrid ARQ protocols for high reliability data transmission on reverse links (uplinks) of power-controlled DS-CDMA personal communication systems have been investigated via analysis and simulations. Convolutional codes (inner codes) were used for error correction only, and shortened BCH codes (outer codes) were employed for error detection. To reduce the feedback channel error, both error-detection and error-correction codes were also used in the negative acknowledgment (NACK) packet. Each NACK contained numbers of all requested data packets, which reduced the number of unnecessary retransmissions. The system's throughput efficiency has been derived as a function of the inner code's frame error rate. The simulation results agreed well with the analysis. Both analysis and simulations demonstrated that the proposed hybrid selective-repeat ARQ protocol was able to ensure virtually error-free data transmission (accepted packet error rate around 10^{-14}) with throughput efficiency around 98% when the required average link SIR was only around 5.1 dB at the Doppler shift of 5 Hz, and about 6.3 dB at the shift of 15 Hz. It was also observed that increasing Doppler shift caused some reduction of the system

throughput when the average received SIR was in the range between 4 and 7 dB. The deleterious effect of NACK message errors on the system's performance was negligible.

A method to determine the optimum target link signal-to-interference ratio per packet transmission maximizing the system's capacity has also been proposed. Its application demonstrated that the number of data users (system capacity) was maximized when the average received SIR per transmission was around 5.1 dB at both 5 Hz and 15 Hz Doppler shifts. The optimum SIR was therefore insensitive to the changes of the Doppler shift. The throughput efficiencies at the optimum SIR and the Doppler shifts of 5 and 15 Hz were equal to 98% and 94%, respectively. Higher values of throughput efficiency per user required higher than optimum average SIR per transmission. An increase of the average SIR beyond its optimum value resulted however in a significant reduction of the overall system capacity.

References

- [1] S. Lin and D.J. Costello, Jr. *Error Control Coding: Fundamental and Applications*. Englewood Cliffs, NJ: Prentice-Hall, 1983.
- [2] S.B. Wicker, *Error Control Systems for Digital Communication and Storage*. Englewood Cliffs, NJ: Prentice-Hall, 1995.
- [3] TIA/EIA/IS-95, "Mobile Station-Base Station Compatibility Standard for Dual-Mode Wideband Spread Spectrum Cellular System," *Telecommunication Industry Association*, July 1993.

- [4] G. Castagnoli, J. Ganz and P. Graber, "Optimum Cyclic Redundancy-Check Codes with 16-Bit Redundancy", *IEEE Trans. Communications*, Vol. 38, No. 1, January 1990, pp. 111-114.
- [5] K. A. Witzke and C. Leung, "A comparison of Some Error Detecting CRC Code Standards", *IEEE Trans. Communications*, Vol. COM-33, No. 9, September 1985, pp. 996-998.
- [6] T. Kasami, T. Klove, and S. Lin, "Linear block codes for error detection", *IEEE Trans. Inform. Theory*, vol. IT-29, Jan. 1983, pp. 131-136.
- [7] J.G. Proakis, *Digital Communications*, 3rd ed., McGraw-Hill, New York, 1995.
- [8] R. Price and P. E. Green Jr., "A communication technique for multipath channels", *Proc. IRE*, vol. 46, March 1958, pp. 555-570.
- [9] L. Jalloul and J. Holtzman, "Performance analysis of DS/CDMA with noncoherent M -ary orthogonal modulation in multipath fading channels", *IEEE Journal on Selected Areas in Communications*, Vol. 12, No. 5, June 1994, pp. 862-870.
- [10] E. K. Hong, K. J. Kim, and K. C. Whang, "Performance evaluation of DS-CDMA system with M -ary orthogonal signaling", *IEEE Trans. On Vehicular Technology*, vol. 45, no. 1, February 1996, pp. 57-62.
- [11] L. F. Chang, F. Ling, D. D. Falconer and N. R. Sollenberger, "Comparison of Two Convolutional Orthogonal Coding Techniques for CDMA Radio Communications Systems", *IEEE Trans. Communications*, Vol. 43, No. 6, June 1995, pp. 2028-2037.
- [12] G. Bao, "Performance evaluation of TCP/RLP protocol stack over CDMA wireless link", *Wireless Networks Journal*, J.C. Baltzer AG, Science Publishers, February 1996, pp. 229-237.

- [13] H. O. Burton and D. D. Sullivan, "Errors and error control", *Proc. IEEE*, 60(11), November 1972, pp. 1293-1310.
- [14] K. Brayer, "Error control techniques using binary symbol burst codes", *IEEE Trans. Comm.*, COM-16, April 1968, pp. 199-214.
- [15] P. S. Sindhu, "Retransmission error control with memory", *IEEE Trans. Comm.*, COM-25, May 1977, pp. 473-479.
- [16] S. Lin and P. S. Yu, "A Hybrid ARQ scheme with parity retransmission for error control of satellite channels", *IEEE Trans. Comm.*, COM-30, July 1982, pp. 1701-1719.
- [17] S. Lin, D. J. Costello, Jr., "Automatic-Repeat-Request error control schemes", *IEEE Communications Magazine*, vol. 22, December 1984, pp. 5-17.
- [18] S. B. Wicker, "Reed-Solomon error control coding for data transmission over Rayleigh fading channels with feedback", *IEEE Trans. On Vehicular Technology*, vol. 41, No. 2, May 1992, pp. 124-133.
- [19] H. Yamamoto and K. Itoh., "Viterbi decoding algorithm for convolutional codes with repeat request.", *IEEE Trans. Information Theory*, IT-26, September 1980, pp. 540-547.
- [20] B.A. Harvey and S.B. Wicker, "Error-trapping Viterbi decoding for type-I hybrid-ARQ protocols", *Canadian Electrical and Computer Engineering Journal*, Vol. 16, No. 1, January 1991, pp. 5-12.
- [21] S. Morgera and V. Oduol, "Soft-decision decoding applied to the generalized type II hybrid ARQ scheme", *IEEE Trans. Comm.*, Vol. 37, April 1989, pp. 393-396.

- [22] S. Kallel and D. Haccoun, "Generalized type II hybrid ARQ scheme using punctured convolutional coding", *IEEE Trans. Comm.*, Vol. 38, No. 11, November 1990, pp. 1938-1946.
- [23] S. B. Wicker and M. D. Bartz, "Type II hybrid ARQ protocols using punctured MDS codes", *IEEE Trans. Comm.*, COM-30, April 1994, pp. 1701-1719.
- [24] TIA/EIA/IS-99, "Data Services Option Standard for Wideband Spread Spectrum Digital Cellular System", *Telecommunication Industry Association*, 1995.
- [25] R. Cam and C. Leung, "Throughput Analysis of Some ARQ Protocols in the Presence of Feedback Errors", *IEEE Trans. Communications*, Vol. 45, No. 1, January 1997, pp. 35-44.
- [26] F. J. MacWilliams, "A theory on the distribution of weights in a systematic code", *Bell Syst. Tech. Journal*, vol. 42, 1963, pp. 79-94.
- [27] C.T. Ong and C. Leung, "On the undetected error probability of triple-error-correcting BCH codes", *IEEE Trans. Information Theory*, Vol. 37, No. 3, May 1991, pp. 673-678.
- [28] A. J. Viterbi, "Error bounds for convolutional codes and an asymptotically optimum decoding algorithm", *IEEE Trans. Information Theory*, IT-13, April 1967, pp. 260-269.
- [29] J. K. Omura, "The Viterbi algorithm", *IEEE Trans. Information Theory*, IT-15, January 1969, pp. 177-179.
- [30] G. D. Forney, "Convolutional codes II: Maximum likelihood decoding", *Information Control*, 25, July 1974, pp. 222-226.
- [31] T. S. Rappaport, *Wireless Communications: Principles & Practice*, Prentice Hall PTR, New Jersey, 1996.

- [32] A. Jalali and P. Mermelstein, "Effects of Diversity, Power Control, and Bandwidth on the Capacity of Microcellular CDMA Systems", *IEEE Journal on Selected Areas in Communications*, Vol. 12, No. 5, June 1994, pp. 952-961.
- [33] A.M. Viterbi and A.J. Viterbi, "Erlang Capacity of a Power Controlled CDMA System", *IEEE Journal on Selected Areas in Communications*, Vol. 11, No. 6, August 1993, pp. 892-900.
- [34] M. L. Honig and J. B. Kim, "Resource Allocation for Packet Data Transmission in DS-CDMA", in *Proc. 1995 Allerton Conference on Communications, Control, and Computing*, Monticello, IL, October 1995, pp. 925-934.
- [35] A. J. Viterbi, A. M. Viterbi, and E. Zehavi, "Performance of power controlled wideband terrestrial digital communications," *IEEE Trans. Commun.*, vol. 41, pp559-569, April 1993.

Chapter 4

CAPACITY ANALYSIS AND CALL ADMISSION CONTROL

4.1 Introduction

Next generation wireless communication systems will be required to provide a variety of services utilizing advanced multiple access techniques. This means that the network must accommodate users with different transmission rates and quality of service (QoS) requirements. As is well known, a DS-CDMA system is interference limited. The Erlang-B formula [1], [2] for computing system capacity when the number of servers (channels) is limited can not be applied here. Capacity analysis of a CDMA system for voice only communications is investigated in [3] and [4], wherein blocking in a voice only DS-CDMA system is defined to occur when the total interference level exceeds a certain threshold value above the background noise level. The maximum Erlang capacity of voice traffic is reached when blocking probability is equal to a given value. Capacity analysis extended to the case of a CDMA system with voice and data users has been presented in [5] and [6]. Nevertheless, when developing call admission policies and a general system capacity formula, a number of issues which have not been yet discussed in the literature have to be taken into account.

A hybrid ARQ scheme can be applied to data transmission as described in Chapter 3. However, the retransmitted packets increase the interference level. To calculate the capacity that a CDMA system can support for data service, the total interference caused

by a data user must be calculated as the normal transmit power multiplied by the average number of transmissions per successfully accepted packet.

A discontinuous transmission scheme is preferred for data traffic of very bursty nature in CDMA systems (see Chapter 2). To rapidly re-establish the physical link after an off-period, an access request message (ARM) is transmitted by a mobile station using step-wise power increases. Discontinuous transmission also requires some overhead, necessary to re-acquire the link after an off-period, which reduces capacity. The advantage in terms of the system capacity of the discontinuous transmission scheme over the continuous one for data traffic of very bursty nature will be justified in this chapter.

Closed loop power control is not feasible for short packet data service. Hence, it is reasonable to assume that the received signal envelope is Rayleigh distributed, and thus the received message power level at the base station is chi-square distributed with a known mean. For long packet data service, closed loop power control can be implemented, but may not be accurate. The received signal-to-interference ratio for each data user can then be approximately modeled as a lognormal variable with a known mean and standard deviation as was done in [3] for voice users.

For high-rate data service, the incoming data stream may be divided into a number of parallel basic-rate streams, and each can then be spread by a combination of its individual orthogonal sequence and a common PN code, in the so-called multi-code CDMA scheme [7]. On one hand this may be a good approach in that the existing IS-95 DS-CDMA

system with minor modifications can support high-rate data service without the problem of very small processing gains for high-rate data users. On the other hand, the multi-code scheme creates rate management problems and does not introduce any overall performance advantage. In CDMA systems, to maintain the same QoS, the higher the data rate, the more the transmitted power is needed, and the interference to other users is larger. When the system resources are limited, admission of a high-rate data service user will degrade the whole system's performance. It is thus essential that a base station identifies each user's data rate as part of its call admission policy, with the objective of efficiently managing system's resources without degrading QoS for both voice and data users.

Capacity analysis of a multi-service CDMA communication system when both continuous and discontinuous transmission schemes are employed, is presented. The system considered can support mixed traffic. The data messages considered in this chapter include long ones for which closed loop power control can be implemented, and short ones for which only open loop power control is feasible. A discontinuous transmission scheme reduces the level of interference and increases system capacity, but also increases the transmission overhead. The effect of this overhead on the system Erlang capacity is discussed in this chapter. For data transmission, when a negative-acknowledgment ARQ protocol is employed, its impact on the system's capacity and the call admission control (CAC) is also investigated. The concept of equivalent bandwidth in CDMA systems is defined to facilitate the development of CAC algorithms. The

proposed CAC algorithms can be employed in a system supporting multi-rate data services. Simulations are conducted to validate the analysis.

4.2 Capacity analysis

4.2.1 Capacity analysis for conventional multiple access systems.

By conventional multiple access systems we mean either FDMA or TDMA systems. In the conventional systems, traffic channels are allocated to users as long as they are available. A channel is an RF frequency band for FDMA, or a time slot and an RF frequency band for TDMA. Incoming traffic is blocked if there is no free channel available. It is a so-called blocked-calls-cleared system. Capacity for a TDMA or FDMA system can be determined using the techniques of queuing theory. The blocked-calls-cleared system can be described as an $M/M/c/c$ queue, where the first M denotes independent (Poisson) arrivals, which implies independent, identically distributed (iid) interarrival times; and the second M indicates that the service times are iid exponential random variables; the first c specifies the number of channels (servers) available; the second c refers to the maximum number of users allowed in the system at any time. Note that in the $M/M/c/c$ queuing system, calls which arrive when there is no channel available are not admitted into the system. The Markov state diagram for this system [2] is shown in Fig. 4.1, where λ is the Poisson arrival rate, and μ the service rate; the states represent the number of users in the system. Define P_j as the stationary state probability for state j ($j=0, 1, \dots, c$). If the system is in the statistical equilibrium, the following equations hold:

$$\lambda P_0 = \mu P_1 \quad (4.1a)$$

$$\lambda P_{j-1} + (j-1)\mu P_{j-1} = \lambda P_{j-2} + j\mu P_j \quad j = 2, \dots, c \quad (4.1b)$$

(4.1a) and (4.1b) lead to the following:

$$P_j = \frac{\lambda}{j\mu} P_{j-1} \quad j = 1, 2, \dots, c \quad (4.2a)$$

$$P_j = \frac{1}{j!} \left(\frac{\lambda}{\mu} \right)^j P_0 \quad j = 0, 1, \dots, c \quad (4.2b)$$

Note that the sum of the state probabilities must be one, that is: $\sum_{j=0}^c P_j = 1$. P_0 is thus

given by:

$$P_0 = \frac{1}{\sum_{j=0}^c (\lambda / \mu)^j / j!} \quad (4.3)$$

It is easy to show that the probability of a call arriving when the system is in state c (probability of blocking) is equal to the probability of the queue being in state c . Hence:

$$P_{\text{blocking}} = P_c = \frac{(\lambda / \mu)^c / c!}{\sum_{j=0}^c (\lambda / \mu)^j / j!} \quad (4.4)$$

where λ / μ is the offered traffic in Erlangs [1], [2]. (4.4) is known as the Erlang B formula.

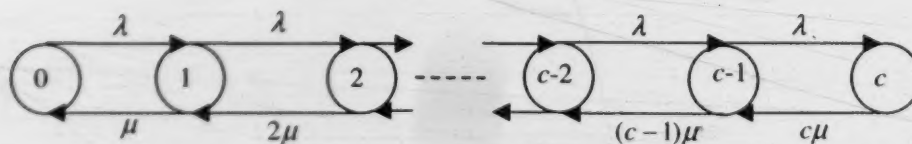


Fig. 4.1. State transition diagram for the $M/M/c/c$ system.

The $M/M/c/c$ queue is a model for a system in which if a call arrives while all servers are busy, the call is immediately cleared with zero holding time and does not reappear during the period of observation (blocked-calls-cleared system). In the $M/M/c$ queueing system shown in Fig. 4.2, however, if all servers are busy when a call arrives, the call will stay in the system (will be queued) until a server becomes available. It is a so-called blocked-calls-delayed system. The analysis for such queueing system involves an extension of the c state Erlang B state transition diagram of Fig. 4.1 to infinity (second $c = \infty$ in $M/M/c/c$). Equation (4.2b) still holds for the first c states. For states beyond c , the steady state probabilities can be obtained from the following recursion:

$$P_j = \frac{1}{c} \frac{\lambda}{\mu} P_{j-1} \quad j = c+1, c+2, \dots, \quad (4.5)$$

which leads to:

$$P_j = \frac{1}{c^{j-c} c!} \left(\frac{\lambda}{\mu} \right)^j P_0 \quad j \geq c \quad (4.6)$$

Since $\sum_{j=0}^{\infty} P_j = 1$, hence, P_0 is given by:

$$P_0 = \frac{1}{\sum_{j=0}^{c-1} \frac{1}{j!} \left(\frac{\lambda}{\mu} \right)^j + \frac{1}{c!} \sum_{j=c}^{\infty} \frac{1}{c^{j-c}} \left(\frac{\lambda}{\mu} \right)^j} = \frac{1}{\frac{(\lambda/\mu)^c}{c!} \frac{c}{c - (\lambda/\mu)} + \sum_{j=0}^{c-1} \frac{1}{j!} \left(\frac{\lambda}{\mu} \right)^j} \quad (4.7)$$

Probability that a newly arrived call is blocked (delayed) is equal to the probability that the system is in a state $N \geq c$:

$$P_{\text{blocking}} = P[N \geq c] = \sum_{j=c}^{\infty} P_j = \frac{1}{c!} \left(\frac{\lambda}{\mu} \right)^c \frac{c}{c - \lambda/\mu} P_0 \quad (4.8)$$

(4.8) is known as the Erlang C formula. We should mention that the Erlang C formula is mainly used in computer networks in which when all servers are busy, new arriving packets (calls) are queued (delayed). The Erlang B formula which is based on the assumption that the blocked calls are cleared immediately is applicable to voice systems (including mobiles).

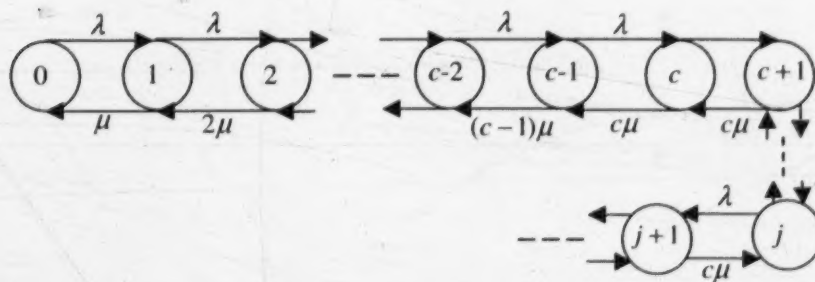


Fig. 4.2. State transition diagram for the $M/M/c$ system.

Now we can examine the system capacity of some popular cellular systems using the Erlang B formula. A total of 50 MHz of RF spectrum [8] is allocated to the AMPS system, in which each user uses two 30 kHz simplex channels to provide full duplex voice and control channels. 7-cell reuse of the total available frequency spectrum is normally employed in the AMPS system. In order to increase system capacity, sectoring which uses directional antennas is also used in the AMPS system. For the case of 120° sectoring, the spectrum available in a cell is broken down into 3 sectored groups. Thus, the 50 MHz bandwidth is divided among 21 sectors (7 cells) into disjoint spectral groups. Thus, the number of channels available per sector can be calculated as follows:

$$c_{\text{AMPS}} = \frac{50 \times 10^3}{2 \times 30 \times 7 \times 3} \approx 39 \text{ channels/sector} \quad (4.9a)$$

For the IS-54 TDMA system [8], three full rate voice users utilize the six time slots in each 30 kHz band. Hence the number of channels (slots) available is just a triple of the AMPS, that is to say, $c_{\text{IS-54}} = 117$ channels/sector (if the same frequency reuse and sectoring patterns are used). The offered traffic capacity per sector for a blocking probability 1% obtained from (4.4) is respectively,

$$(\lambda/\mu)_{\text{AMPS}} = 28.1 \text{ Erlangs} \quad (4.9b)$$

$$(\lambda/\mu)_{\text{IS-54}} = 100.1 \text{ Erlangs} \quad (4.9c)$$

Suppose that each user makes 1 call per hour at an average call duration of 3 minutes. Then the total number of users that can be supported per sector at 1% blocking is equal to (AMPS and IS-54, respectively):

$$K_{\text{AMPS}} = \frac{28.1}{3/60} = 562 \text{ users}$$

$$K_{\text{IS-54}} = \frac{100.1}{3/60} = 2002 \text{ users}$$

We see that the digital cellular system (IS-54) can offer much higher capacity than the corresponding analog system (AMPS).

4.2.2 Capacity analysis for entirely interference limited systems.

4.2.2.1 Intuitive CDMA capacity analysis.

For the interference limited CDMA system, the capacity analysis is much different from the conventional server limited system, since there is no hard limit on the total number of

the available channels. In the ideal case, if all users' PN codes are mutually orthogonal, which implies no mutual interference among users, the system can support infinite number of subscribers as long as there are enough processors available to serve all users. This will of course never happen in the real world. Practically, there are non-zero cross-correlations and auto correlations between PN waveforms. The noise to a desired user includes not only the background noise, but also the interference from all other users in the whole network. The capacity analysis of a CDMA system is based on the analysis of total received interference rather than the total available bandwidth.

Before we go into more detailed study of the CDMA capacity, we will present a more intuitive approach to determining it.

To develop the equation for the CDMA system capacity, we start with the equation for S/S_I , where S is the received signal power from a mobile station in a cell, and S_I is the total received interference plus noise power. The S/S_I ratio is:

$$\frac{S}{S_I} = \frac{rE_b}{WI_0} \quad (4.10)$$

where r is transmission bit rate, and all other symbols were explained in (3.31) of Chapter 3. Assuming perfect power control, the interference power S_I is equal to $(K-1)S$, where K is the number of users transmitting in the band W . Note that this assumes that the background noise is negligible as compared to the multiuser interference. Thus S/S_I is equal to $1/(K-1)$. This results in the following equation for

the CDMA capacity in a non-cellular environment (no neighboring interference) without additional system features.

$$K - 1 = \frac{W/r}{E_b/I_0} \quad (4.11)$$

where W/r is the CDMA system processing gain.

Two more features can be added to the CDMA system in order to diminish interference, namely voice activity detection and cell sectoring. In a typical full duplex two-way conversation, numerous measurements [9] have established that the duty cycle of each voice call is usually about three-eighths. In a CDMA system, transmission rate is reduced when there is no speech signal to transmit. By doing this, the interference to other users is substantially reduced. The capacity is thus increased proportionally to the overall rate reduction. We denote this factor as the voice activity gain, G_v , and $G_v \approx 2.67$ (8/3). As mentioned before, cell sectoring using directional cell site antennas can also be used to increase capacity. For the case of the typical 120° sector antennas, the interference is only received in the direction of one-third of the mobile stations, or the total received interference is reduced by a factor of 3. Thus the system capacity is almost tripled. Denote this factor as the antenna gain, G_A . Accounting for side lobes, G_A is equal to approximately 2.4 [10].

Finally, for a cellular system in which all users share the same frequency bandwidth of W , the actual frequency reuse factor is not one. We must consider the interference

introduced not only by users within a given cell, but also by all users in all other cells.

Thus, we need to consider the factor which is defined as:

$$f = \frac{\text{interference from other cells}}{\text{interference from given cell}} \quad (4.12)$$

This value varies depending on propagation and several other factors. Consequently, the capacity formula in (4.11) must be reduced by the factor of $1 + f$. Overall, considering these three factors discussed above, the CDMA system capacity formula in (4.11) can be expressed as:

$$K = \frac{W/r \cdot G_v G_A}{E_b/I_0 (1+f)} \quad (4.13)$$

In the IS-95 system [11], all users share 1.25 MHz bandwidth, and the system processing gain is 128. Simulation and experiments reported in [3], [4], [10], [12] suggest that it is possible to achieve high quality voice communications with $E_b/I_0 = 6 \sim 7$ dB. Thus, the CDMA system using only 1.25 MHz can support ($E_b/I_0 = 7$ dB or 5 on a linear scale, is used):

$$K = \frac{128}{5} \times \frac{2.67 \times 2.4}{1+0.55} \approx 105 \text{ channels/sector} \quad (4.14)$$

where $f = 55\%$ is used [3], [4], [10]. By using (4.4), with 1% blocking, Erlang capacity per sector is about 88.8 Erlangs in 1.25 MHz. Recall that in the AMPS and IS-54, even though the entire operating bandwidth (50 MHz for both reverse and forward frequency bands) is used, the Erlang capacities per sector are only 28.1 Erlangs for the AMPS (4.9b) and 100.1 Erlangs for the IS-54 (4.9c). A huge capacity increase is achieved by the CDMA system.

We note that the number in (4.14) is only result of intuitive analysis. In the actual IS-95 system [11], the forward link capacity is limited by the number of available Walsh codes. For example, on the forward link of the IS-95 system, only 64 Walsh codes or 64 channels are used. However, in order to work properly, the system needs one pilot channel, one synchronization channel, and at least one paging channel [11]. In other words, at most 61 channels per sector can be used to carry traffic. With 1% blocking, this gives 47.9 Erlangs per sector. This is the maximum capacity that the IS-95 system can provide. Nevertheless, it is still much larger than that provided by AMPS or IS-54 systems.

4.2.2.2 Detailed capacity analysis

The theoretical criterion used in determining the CDMA system capacity states that the total received noise-plus-interference power to the background noise can not exceed a given threshold [3], [4], [10]. The condition for blocking used in (3.31) results from that criterion. In the following analysis, when a user needs high-rate data service and the data rate is the N -th multiple of a basic-rate (in IS-95, the basic-rate is 9.6 kbps), the data stream is converted into N parallel basic-rate streams. When the transmission rate is not an integer multiple of the basic rate, for example, $r_h = Nr + r_l$, where r_h , r , and r_l are high, basic, and low data rates, respectively, and $r_l < r$, the user's data is converted into one low-rate plus N basic-rate parallel streams [7]. Thus in the following analysis, all transmission rates are either less than or equal to the basic-rate.

For a multi-service DS-CDMA network, blocking is defined to occur when

$$\sum_{i=1}^{K_v} v_i E_{b,vi} r + \sum_{i=1}^{K_l} l_i (1+h) E_{b,li} \zeta_{li} r + \sum_{i=1}^{K_s} s_i (1+h) E_{b,si} \zeta_{si} r + N_0 W > I_0 W \quad (4.15)$$

where r is the basic data rate, K_v , K_l , and K_s are numbers of voice, long-packet data, and short-packet data service users per cell, respectively, $E_{b,vi}$, $E_{b,li}$, and $E_{b,si}$ are bit energies for voice, long-packet data and short-packet data users, h is the overhead interference contributed by synchronizing messages used to re-establish physical link when a discontinuous transmission scheme is adopted [13] (also see Chapter 2), ζ_{li} and ζ_{si} are factors obtained by dividing long-packet and short-packet data rates by the basic rate r , and v_i , l_i , and s_i are activity factors of voice, long-packet and short-packet data users, defined as follows:

$$v_i = \begin{cases} 1 & \text{with probability } p_v \\ \alpha_v & \text{with probability } 1 - p_v \end{cases} \quad (4.16)$$

$$l_i = \begin{cases} 1 & \text{with probability } p_l \\ \alpha_l & \text{with probability } 1 - p_l \end{cases} \quad (4.17)$$

$$s_i = \begin{cases} 1 & \text{with probability } p_s \\ \alpha_s & \text{with probability } 1 - p_s \end{cases} \quad (4.18)$$

p_v is the probability of a voice user being active; p_l and p_s are the probabilities of long-packet and short-packet data users transmitting at the full bit rate; α_v , α_l , and α_s are relative low transmission bit rates transmitted during idle periods to maintain synchronization. We note that for the discontinuous transmission scheme, $\alpha_l = 0$, and for continuous transmission scheme, $h = 0$. Only discontinuous transmission of short packets has been considered in this work ($\alpha_s = 0$). Note that the bit energies in (4.15) for data users are the effective ones including the effect of retransmissions (see (3.31)).

Inequality (4.15) can be presented as follows:

$$Z = \sum_{i=1}^{K_v} v_i \epsilon_{vi} + \sum_{i=1}^{K_l} l_i (1+h) \zeta_{li} \epsilon_{li} + \sum_{i=1}^{K_s} s_i (1+h) \zeta_{si} \epsilon_{si} > (W/r)(1-\xi) \quad (4.19)$$

where ϵ_{vi} , ϵ_{li} and ϵ_{si} are bit energy-to-interference ratios defined as $\epsilon_{vi} = E_{b,vi} / I_0$, $\epsilon_{li} = E_{b,li} / I_0$, and $\epsilon_{si} = E_{b,si} / I_0$, respectively.

To facilitate further analysis, the following additional assumptions have been made:

- The numbers of voice, long-packet data and short-packet data calls, K_v , K_l , and K_s , respectively, are Poisson random variables with means λ_v / μ_v , λ_l / μ_l , and λ_s / μ_s , where λ_v refers to the number of service requests per second (Poisson arrival rate), and $1/\mu_v$ is the mean service or call holding time in seconds per call.
- The received $SIR = E_b / I_0$ for each voice user can be modeled as a lognormal random variable with a known mean m_v dB and a standard deviation σ_v dB [3].

The following comments can also be made in relation to (4.19):

- In a CDMA system using an ARQ protocol, there is an optimum SIR which minimizes the resulting interference and naturally leads to increased capacity for an interference-limited system; this issue is addressed in [14] and Chapter 3. The optimum SIR is still a lognormal random variable with mean m_l dB and deviation σ_l dB.
- For a data user requesting short message service, there is also an optimum SIR if an ARQ protocol is used. Due to the absence of closed-loop power control, the SIR ϵ_{si} in (4.19) is a chi-square random variable with two degrees of freedom (or an

exponential random variable). It can be written as $\varepsilon_{si} = z_i \bar{E}_{b,s} / I_0$, where $\bar{E}_{b,s} / I_0$ is the average bit energy-to-interference ratio, and z_i is an exponential variable with unity standard deviation; $p(z_i) = e^{-z_i}$.

The Erlang capacity of a multi-service CDMA network is established by setting the blocking probability equal to a given value ϕ (typically $\phi = 1\%$) [3]:

$$\Pr[Z > (W/R)(1 - \xi)] = \phi \quad (4.20)$$

One simple approach to compute the Erlang capacity, which was also used in [3], is to assume that Z is Gaussian. With this assumption (4.20) can be expressed as follows:

$$Q\left(\frac{(W/R)(1 - \xi) - E(Z)}{\sqrt{\text{Var}(Z)}}\right) = \phi \quad (4.21)$$

where $Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-t^2/2} dt$. If we take into account interference from other cells, and assume that all cells are equally loaded with traffic, a derivation similar to that in [3] will yield the mean and variance of Z :

$$\begin{aligned} E(Z) = & \frac{\lambda_v}{\mu_v} (1 + f) [p_v + \alpha_v (1 - p_v)] \exp(\beta m_v + (\beta \sigma_v)^2 / 2) \\ & + \frac{\lambda_l}{\mu_l} (1 + f) (1 + h) [p_l + \alpha_l (1 - p_l)] E[\zeta_l] \exp(\beta m_l + (\beta \sigma_l)^2 / 2) \\ & + \frac{\lambda_s}{\mu_s} (1 + f) (1 + h) [p_s + \alpha_s (1 - p_s)] E[\zeta_s] \frac{\bar{E}_{b,s}}{I_0} \end{aligned} \quad (4.22)$$

$$\text{Var}(Z) = \frac{\lambda_v}{\mu_v} (1 + f) [p_v + \alpha_v^2 (1 - p_v)] \exp(2\beta m_v + 2\beta^2 \sigma_v^2)$$

$$\begin{aligned}
& + \frac{\lambda_l}{\mu_l} (1+f)(1+h)^2 [p_l + \alpha_l^2 (1-p_l)] E[\zeta_l^2] \exp(2\beta m_l + 2\beta^2 \sigma_l^2) \\
& + 2 \left(\frac{\bar{E}_{b,s}}{I_0} \right)^2 \frac{\lambda_s}{\mu_s} (1+f)(1+h)^2 [p_s + \alpha_s^2 (1-p_s)] E[\zeta_s^2]
\end{aligned} \tag{4.23}$$

where f is defined in (4.12), $E[\zeta_l]$ and $E[\zeta_s]$ are the first moments of ζ_{li} and ζ_{si} , $E[\zeta_l^2]$ and $E[\zeta_s^2]$ are the second moments, respectively, and $\beta = \ln(10)/10$.

Estimates of the number of simultaneous long and short packet data users, as well as voice users, that a CDMA system with a processing gain of 256 can support under the prescribed QoS requirement of 1 percent outage probability are shown in Fig. 4.3 and Fig. 4.4; $E[K_v]$ is the average number of voice users in the system, *CT* and *DT* stand for continuous and discontinuous data transmission, respectively. The parameters used to obtain the numerical results are listed in Table 4.1. Note that the overhead for discontinuous schemes is assumed 10% which is quite large, since as demonstrated in chapter 2 when the SIR is around 12 dB, one ARM is enough for coarse synchronization, and the length of an ARM is only about 4~5 thousand PN chips.

Table 4.1. Parameters used to obtain the numerical results of the Erlang capacities

Symbol	Value
m_v, m_l	6 dB
σ_v, σ_l	2.5 dB
$\bar{E}_{b,s} / I_0$	7 dB

α_v	0.125
α_l	0.125 (CT) or 0 (DT)
α_s	0
p_v	0.4
p_l	0.2
p_s	0.2
$\eta = N_0 / I_0$	0.25
h	0.1 (DT) or 0 (CT)
f	0.55

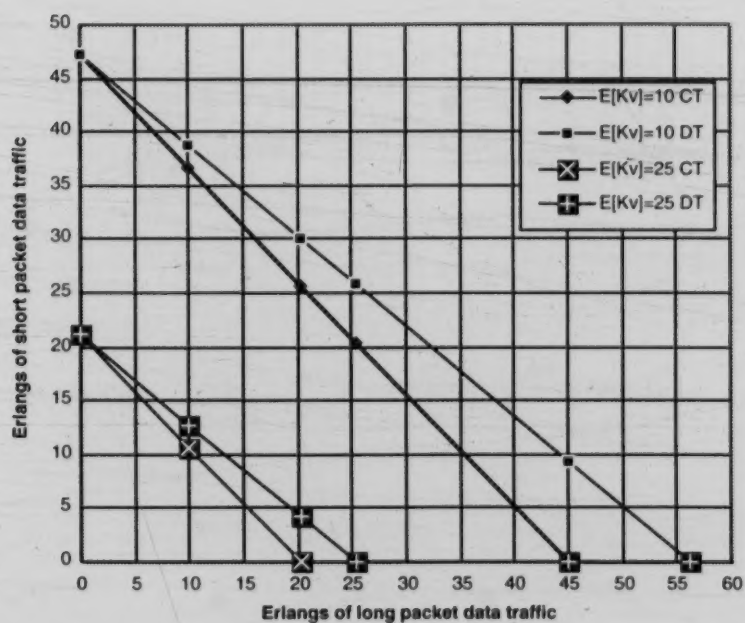


Fig. 4.3. Erlang capacities for voice users, short packet data , and bursty ($p_l = 0.2$) long packet data users.

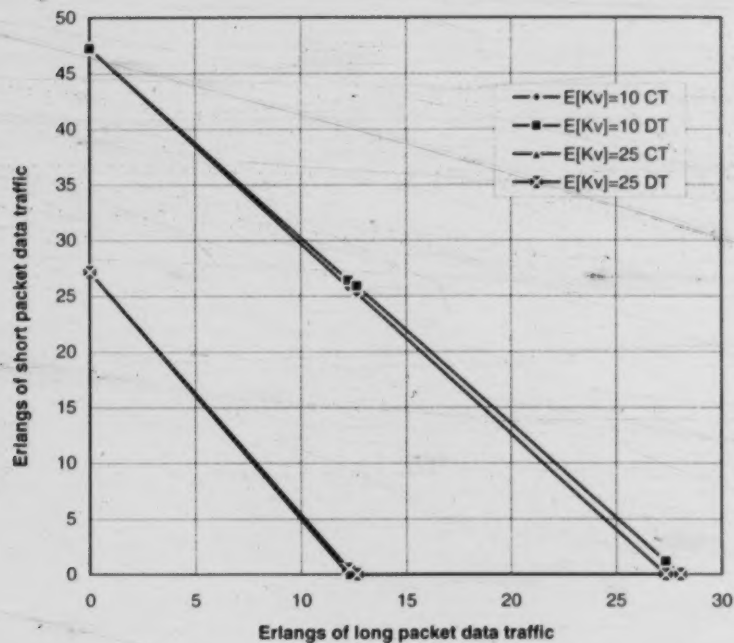


Fig. 4.4. Erlang capacities for voice users, short packet data, and regular ($p_l = 0.4$) long packet data users.

It can be observed from Fig. 4.3 that the discontinuous scheme can provide higher capacity for interference-limited CDMA systems than the continuous scheme even if the overhead is assumed to add 10 percent to the interference caused by a bursty user. However, when the data traffic is regular, the advantage of the discontinuous transmission scheme is trivial as shown in Fig. 4.4.

Usually Erlang capacity is not a good measure for packet data. However, on the reverse link of the CDMA system no mechanism is assumed here to delay transmission of packets and a call that is admitted is allowed to proceed immediately. Thus a packet data

call will be a circuit-switched call as far as radio link is concerned, and the use of Erlang capacity to evaluate packet data in this case is reasonable.

4.3 Call admission policies

The fundamental call admission control criterion requires that the ratio of the total measured received power to the background noise at the base station is not allowed to exceed a given fixed threshold [3], [10]. In general, CDMA systems are primarily interference limited, and all cells share the same bandwidth. A newly accepted user will not only introduce interference to its own cell, but also to adjacent, and even remote cells. Another issue that needs to be addressed is that, in practice, there is a propagation and processing delay before a newly accepted data packet reaches the receiving end, and after that time, the total interference introduced by all active users will be different from what it was when the new data user was admitted, since the received SIR is a random process. This delay is denoted by T . The effect of each call request on adjacent cells is considered in [15], but the change of interference after time T is not taken into account. As will be demonstrated below, the algorithm in [15] is only applicable to the voice only communication system. A good analysis framework for CAC schemes considering future traffic and handoff is given in [16], and an overall performance comparison of CAC algorithms for a cellular system that carries a single class of homogeneous traffic and uses fixed channel assignment is investigated in [17]. However, neither of them is suitable for CDMA systems that support multi-rate services, since the algorithms in [16] and [17] apply only to strictly channelised systems (like TDMA or FDMA). In this section, the SIR-based CAC algorithm [15] is reviewed, and then a general CAC criterion

is developed which will serve as a framework for development of some practically feasible CAC algorithms. Two CAC schemes will then be introduced. One of them uses prediction to account for the changed interference after the delay T .

4.3.1 SIR-based CAC algorithm

Two SIR-based CAC were proposed in [15], namely the local and global algorithms. The first algorithm is solely based on the SIR measurement at the local cell base station. Define the system residual capacity as χ_i for cell i . The algorithm used in [15] can be simply described as follows:

- The base station in each cell i makes periodic measurement of its reverse link SIR_i ;
- If SIR_i is large enough so that $\left[\frac{1}{SIR_{th}} - \frac{1}{SIR_i} \right] > 0$, where SIR_{th} is called the SIR threshold at the base station, and $\lfloor x \rfloor$ denotes the largest integer less than or equal to x , then $\chi_i > 0$;
- For each new call in cell i , if $\chi_i > 0$, the new call is accepted, and the residual capacity is reduced by one; otherwise, the call is rejected.

This algorithm can not work in the cellular environment, since it does not take into account the interference introduced by a new user to its neighboring cells. For example, suppose cell-1 and cell-2 are neighbors, and the measured SIR in cell-2 is close to its threshold value (no more users can be admitted into cell-2), while the measured SIR in cell-1 is much higher than the threshold value. Thus, one or even more users can be

admitted into cell-1 according to the algorithm. Obviously when one or more users admit into cell-1, cell-2 is overloaded. In order to make this algorithm work in the cellular environment, a global algorithm was proposed in [15]. It is similar to the local algorithm. The only difference is the estimation of the residual capacity, which is updated according to the following formula:

$$\chi_i = \begin{cases} \min\{\chi'_i\} & \text{if } \min\{\chi'_i\} > 0 \\ 0 & \text{otherwise} \end{cases} \quad (4.24)$$

where for a network with N cells, $j = 1, 2, \dots, N$, and

$$\chi'_i = \begin{cases} \left[\frac{1}{\text{SIR}_{\text{TH}}} - \frac{1}{\text{SIR}_i} \right] & \text{if } i = j \\ \left[\frac{1}{\beta} \left(\frac{1}{\text{SIR}_{\text{TH}}} - \frac{1}{\text{SIR}_i} \right) \right] & \text{if } i \neq j \end{cases} \quad (4.25)$$

where β is called the adjacent cell interference coupling coefficient.

This algorithm takes into account the interference introduced by a new user to the whole network. However, this algorithm can probably be used only for admitting voice users, and it can not be used to accept high data rate users. For example, as shown in Fig. 4.5, since the measured SIR is always higher than the threshold SIR, a requested call may be admitted into the network according to the algorithm. For a voice call, this is fine, since a voice call may only reduce the measured SIR a little, and after the new voice user has been admitted into the cell, the measured SIR may still be higher than the SIR_{TH} , so the quality of communications is still maintained for all users in the cell. However, when a high data rate user enters the cell, the introduced interference could be orders of magnitude greater than that of a voice user. The next generation CDMA systems are supposed to support a data rate up to 2 Mbps [18], and the data rate for voice services is

only 9.6 kbps. Even with the same QoS requirement, such as 1% of the frame error rate, the interference contributed by a data user with a rate of 2Mbps is roughly equal to the interference contributed by 200 voice users! In other words, when a high rate data user is admitted into the cell using the SIR-based algorithm, the measured SIR may drop well below the threshold SIR in Fig. 4.5, and the QoS for all users in the cell will be severely degraded.

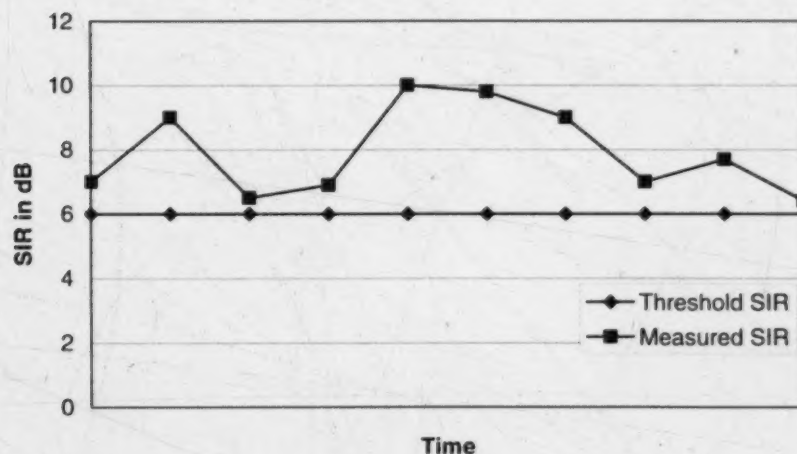


Fig. 4.5. Illustration of the SIR-based CAC algorithm.

4.3.2 General call admission control (CAC) criterion

The SIR-based CAC algorithm works well for voice or low rate data services, because when a voice user requests a service, the network implicitly knows the amount of interference that a voice user may contribute. This is based on the fact that normally the data rate and the required minimum SIR for voice service are known in advance. For a

CAC algorithm that can admit a high rate data user, the key issue is to consider the potential large interference contributed by a high rate data user.

In order to avoid blocking, an equation similar to (4.19) accounting for both voice and data users can be written as follows:

$$\sum_{j=1}^{K_v} \epsilon_{vj} r_{vj} + \sum_{j=1}^{K_d} N_r(\epsilon_{dj}) \epsilon_{dj} r_{dj} + \sum_{o \text{ other cells}} N_r(\epsilon_{j(o)}) \epsilon_{j(o)} r_{j(o)} \leq W(1-\xi) \quad (4.26)$$

where indexes v and d refer to voice and data (long and short) users respectively, $N_r(\epsilon_{dj})$ is the average number of times that a data message must be transmitted before it is accepted, and it is a function of the SIR ϵ_{dj} . Obviously for voice users $N_r(\epsilon_{vj}) = 1$.

(4.26) is conservative, because it includes the non-existent interference term caused by the desired user. Note that, in order to simplify the analysis below, the activity factors for both voice and data services have been assumed to be one. The numerical results presented later in section 4.3.5 consider the actual activity factors.

To simplify the analysis, we express (4.26) as

$$\sum_{j=1}^{K_v} W_{vj} + \sum_{j=1}^{K_d} W_{dj} + \sum_{o \text{ other cells}} \sum_{j=1}^{K_o} W_{j(o)} \leq W_T \quad (4.27)$$

In the above expression, $W_T = W(1-\xi)$, $W_{vj} = \epsilon_{vj} r_{vj}$, and $W_{dj} = N_r(\epsilon_{dj}) \epsilon_{dj} r_{dj}$. We shall henceforth call W_T the system equivalent bandwidth, and W_{dj} and W_{vj} equivalent bandwidths of the j -th data and voice users. $W_{j(o)}$ is the equivalent bandwidth of the j -th users from other cells. A user's equivalent bandwidth is defined as its equivalent SIR times its relative bit rate. Note that the equivalent SIR for a data user is the actual link

SIR times the average number of times that a data packet must be transmitted. This equivalent SIR accounts for the impact of ARQ on the system, and results from an observation that, when accepting a data packet for transmission, the system has to allocate to it a resource which will guarantee it is successfully delivered to its sink without rejecting or buffering it during its transmission. As proposed in [14] and Chapter 3, the retransmission scheme uses a continuous selective-repeat ARQ, which keeps transmitting admitted user's packets until they are successfully delivered.

If (4.27) is violated, the system's performance degrades. For data packets, due to the error detection capability at the receiving end, more interference leads to more retransmissions, and the packet error rate can still be maintained. For voice users, however, since retransmission is not allowed, their quality of service (QoS) deteriorates. Thus each base station has to control all users active in its cell. As stated in section 4.1, when a data user requests admission, a base station identifies the user's desired bit rate and also its required QoS, which can be mapped to its equivalent SIR. As shown in Fig. 4.6, when a user in cell C_0 requests admission to the network at time t , its interference to the entire network has to be considered. In other words, for a given blocking probability ϕ , only when the following inequality holds for all 19 cells, the request is approved:

$$\Pr[(\sum_{j=1}^{K_{C_i}} W_j + \sum_{\substack{u=0 \\ u \neq i}}^{18} \sum_{j=1}^{K_{C_u}} W_{j(u)}) \geq W_T^{C_i}] \leq \phi, i = 0, 1, 2, \dots, 18 \quad (4.28)$$

where K_{C_i} and K_{C_u} are numbers of users in cells C_i and C_u .

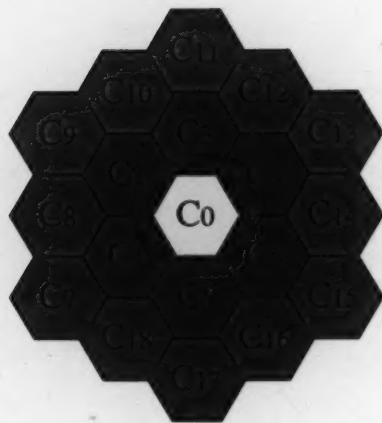


Fig. 4.6. Cell layout.

If the effect of traffic variations during the interval T is taken into account, the request is approved only if the following modified inequality holds at time $t+T$.

$$\Pr[(\sum_{j=1}^{\tilde{K}_{C_0}} \tilde{W}_j + \sum_{\substack{u=0 \\ u \neq i}}^{18} \sum_{j=1}^{\tilde{K}_{C_u}} \tilde{W}_{j(u)}) \geq W_T^{C_i}] \leq \phi \quad (4.29)$$

where $\tilde{W}$ is the user's predicted equivalent bandwidth. The number of users $\tilde{K}_{C_i}$ and $\tilde{K}_{C_u}$ may be different from those in (4.28) because of the possibilities of handoffs and call termination.

It is complicated to implement a CAC algorithm based on either (4.28) or (4.29) in a cellular system. In case of (4.28), the problem is how to determine the interference from other cells, especially for nonuniformly distributed traffic. For example in Fig. 4.6, if cell C_{11} is already at the limit of its capacity, a newly admitted user in C_0 will cause users' QoS in C_{11} to drop below the acceptable level. Hence, distributed CAC is necessary [16].

The dynamic condition (4.29) additionally accounts for the possibility of call handoffs, terminations and changing propagation conditions during the delay interval T .

Obviously it is not an easy task to develop a good algorithm which can overcome these difficulties. Our objective is to develop an algorithm which can be implemented easily in a system that supports multi-rate services and also increases the system capacity in the average sense, and not at some particular instance of time.

4.3.3 Non-predictive CAC (NPCAC).

From (4.27), we see that the system bandwidth W_T may be occupied by users in the home, neighboring and remote cells. We split W_T into two parts, of which ρW_T is used to accommodate users in the home cell, $(1 - \rho)W_T$ is reserved for users in all other cells. ρ is actually the CDMA system's frequency reuse efficiency which is defined as the ratio of interference from mobile stations within a given cell to the total interference from all cells. It can be obtained by simulation, or by real time traffic measurements. It depends on traffic loading in each cell, propagation environment and path loss exponent between cells. Some examples on how to determine ρ will be addressed in section 4.3.5. One simple CAC algorithm is proposed in the following.

Let the total equivalent bandwidth occupied by all active users in cell C_0 in Fig. 4.6 be W_a , and the incoming user's required equivalent SIR be ϵ_d , bit rate r_d , and equivalent bandwidth $W_d = \epsilon_d r_d$. The newly accepted user in cell C_0 will introduce interference to

cells $C_1 - C_{18}$, and therefore every cell needs to check if there is enough bandwidth to accommodate the new user. Denote the equivalent interference to cell C_i ($i=1, \dots, 18$) contributed by the new user as $\varphi_i W_d$. Depending on the distance between cells, propagation environment and path loss exponent, the factor φ_i varies significantly. For example, according to our simulation results, the average value of φ_i is around 8% for $i=1, \dots, 6$, and 0.25% for $i=7, \dots, 18$. The access request is approved by the base station, if the following conditions are satisfied:

$$(1) W_d \leq \rho W_T - W_a. \quad (4.30)$$

(2) $\varphi_i W_d \leq REB_i$ for cells C_i ($i=1, 2, \dots, 18$); REB_i is the cell's remaining equivalent bandwidth.

W_d and φ_i are random variables. Exceeding thresholds given above is tolerable as long as the probability of such events is smaller than the design value of the blocking probability.

4.3.4 Predictive CAC (PCAC).

This algorithm attempts to take into account future traffic variations due to handoffs and call departures. Similarly to the NPCAC scheme, new call requests to enter cell C_0 requiring bandwidth W_d can be accepted if the following conditions are satisfied:

$$(1) W_d + \sum_{j=1}^{\lambda_h T} W_{hj}^0 + \sum_{j=1}^{\tilde{K}_{C_0}} \tilde{W}_{aj}^0 P_{aj}^0 \leq \rho W_T \quad (4.31)$$

where λ_h represents the average call handoff rate, W_{hj}^i is the equivalent bandwidth of user j handoffed to cell C_i , $\tilde{W}_{aj}^i$ is the predicted value of the active user j 's equivalent

bandwidth, P_{aj}^t is the probability of user j remaining in cell C_i during time T , and $\tilde{K}_{C_0}$ is the number of active users in cell C_i . The second term in (4.31) represents equivalent bandwidth consumed by users arriving in the home cell due to handoffs during the delay interval T . The third term in (4.31) is the equivalent bandwidth taken by users already present in the home cell; it is evaluated after the delay T and takes into account call departures. Normally call duration is exponentially distributed with mean holding time $1/\mu_j$, and hence $P_{aj} = e^{-\mu_j T}$.

- (2) $\phi_i W_d \leq REB_i$ for cells C_i ($i=1,2,\dots,18$); REB_i is the cell's remaining equivalent bandwidth.

Although condition (2) seems to be the same as the corresponding one in NPCAC, the difference is that when calculating REB_i , the cell C_i also has to predict its future traffic and consider newly handoffed, and terminated calls.

A linear predictor of order L is used to predict the future traffic from its present and past values. The prediction equation is the following:

$$\tilde{W}_a(j) = \sum_{k=1}^L a_k W_a(j-k) \quad (4.32)$$

where $\tilde{W}_a(j)$ is the predicted occupied equivalent bandwidth at time $t = jT$ determined from occupied equivalent bandwidths $W_a(j-k)$ measured at times $t = (j-k)T$. Here we have assumed there are L available samples. $\{a_k\}$ are the non-adaptive predictor coefficients determined using the minimum mean squared error (MMSE) criterion.

It should be mentioned that neither NPCAC nor PCAC algorithm allows home cell users to occupy the bandwidth reserved for other cell users. Often the interference to the local cell contributed by several users from other cells is equivalent to the interference generated by only one local user. Hence, allowing a local user to occupy bandwidth reserved for other cell users could cause severe reduction of the system capacity (one admitted local user prevents admission of several users in other cells). This is similar to giving priority to users in neighboring cells as discussed in [19].

4.3.5 Simulation and numerical results

A system containing 19 cells as shown in Fig. 4.6, is considered. The propagation loss is assumed to be a product of the 4th power of distance and a log-normal component due to shadow fading. Hence, the attenuation for a user at a distance r from a base station is given by:

$$\alpha(r, \zeta) = r^4 10^{\zeta/10} \quad (4.33)$$

where ζ is a Gaussian random variable with zero mean and standard deviation of 8 dB.

We use the same simulation model as in [12]. The fast power control algorithm involves signal to interference ratio measurements at the 1.25 millisecond intervals. The parameters used in the simulation and numerical results are given in Table 4.2.

Table 4.2. Parameters used in the simulation and numerical results for CAC algorithms

Parameter	Value
System bandwidth	5 MHz
Processing gain	256
Number of discernible propagation paths	6
Power profile	0, -3, -6, -9, -12, -18 dB
Power control step	1 dB
SIR measurement interval	1.25 ms
Average number of voice users	25
η	0.25
p_v	0.4
p_l	0.2
$1/\mu$	3 minutes
$1/\lambda_h$	100 s
α_v, α_l	0.125

As stated in Section 4.3.4, a linear predictor is used to predict future traffic. Since the equivalent bandwidth is defined as effective SIR times bit rate, and bit rate is assumed constant, the predictor actually predicts the future SIR. Sample SIRs obtained by simulation and prediction are presented in Fig. 4.7; the number of predictor taps is 4. It has been found that the predictor performs best when the number of taps is in the range of 3 to 8.

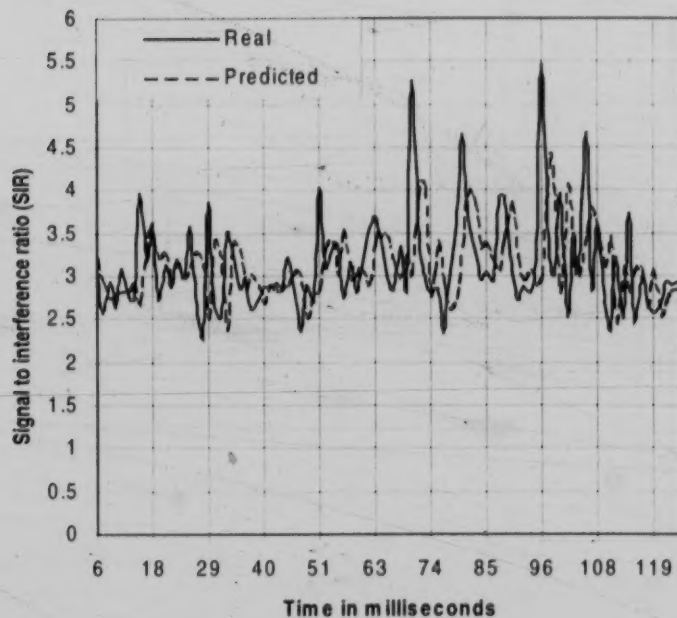


Fig. 4.7. Samples of actual and predicted SIR.

The design parameter ρ can be obtained by simulation or by real time traffic measurements, as mentioned before. Cells $C_1 - C_6$ each contribute about 8.1% of the total interference to cell C_0 , and cells $C_7 - C_{18}$ about 0.25% each. When the average traffic in cells $C_0 - C_{18}$ is the same, the value of ρ

is $\rho = \frac{1}{1 + 6 \times 0.081 + 12 \times 0.0025} = 0.66$. Increasing ρ will increase the local cell's

capacity, but the capacity in adjacent cells will be reduced. To illustrate this, the available equivalent bandwidths (AEB) in Mbps for cells $C_0 - C_{18}$ are given in Table 4.3 for different values of ρ . Obviously for equal traffic loading in cells $C_0 - C_{18}$, the whole network's capacity for $\rho = 0.66$ is much greater than that at $\rho = 0.866$ and unequal traffic

loads in different cells. Although the capacity of cell C_0 is increased for $\rho = 0.866$, the total available bandwidths in cells $C_1 - C_{18}$ are reduced. This is easily explained by the fact that the cell C_0 only reserves $(1-0.886)W_T$ for interference from other cells. However for unequal cell loading, a heavily loaded cell can reserve less bandwidth for out-of-cell interference from lightly loaded neighboring cells, and then the capacity of the hot cell can be increased. This can be observed in Table 4.3 for lightly loaded $C_1 - C_{18}$.

Table 4.3. Bandwidth available for home and remote cells corresponding to different values of ρ

	AEB ($\rho = 0.66$)	AEB ($\rho = 0.866$)
C_0	1.825	2.394
C_{1-18}	32.85	9.846

The effect of the ARQ on the system's performance is shown in Fig. 4.8. Continuous selective-repeat ARQ schemes will generate more interference than a transmission scheme without ARQ, and thus increase the system's blocking probability. For example in Fig. 4.8, if the effect of ARQ is not considered, and 66 data users are admitted into the cell for the desired blocking probability of 2%, the actual blocking probability is around 2.3%. Therefore, the cell can not handle the total 66 data users at 2% blocking. This is why we define the data user's equivalent bandwidth as its effective cumulative SIR (not the actual link SIR) times the bit rate.

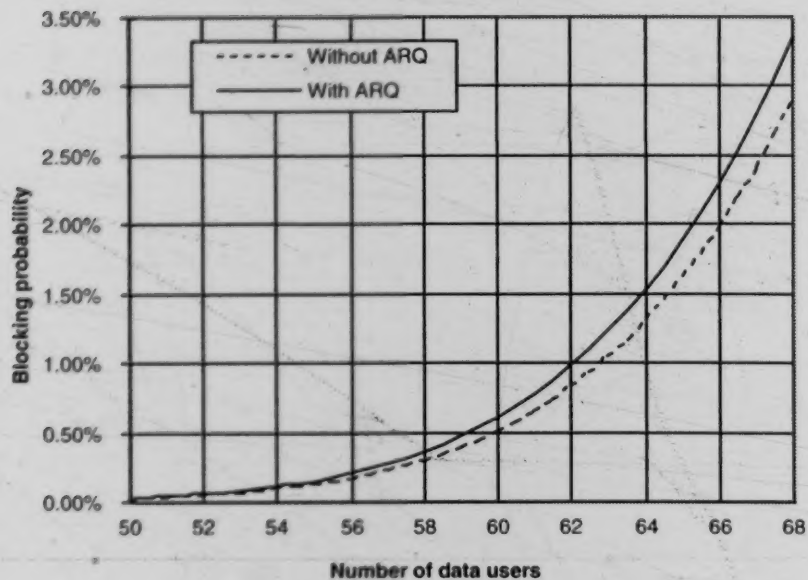


Fig. 4.8. Impact of the ARQ on the system's performance.

Finally, the performance comparison of NPCAC and PCAC algorithms is presented in Fig. 4.9. It can be seen that if the delay T is very short, the advantage of PCAC algorithm is negligible. However, if T is very long, PCAC clearly outperforms NPCAC. The curves have been obtained by assuming that traffic is predictable, and that parameters such as handoff rate and call duration distribution are available. Actual traffic measurements and statistical analysis of the measured data are necessary to obtain these parameters.

Performance measures such as handoff dropping probability and average packet delay are not included in our CAC algorithms. This is justified by the fact that in a CDMA system, as long as the base station can provide enough PN despreaders, call dropping does not

occur. As far as the average packet delay is concerned, it depends on the number of retransmissions and the average queuing delay. Our algorithms require the base station to allocate bandwidth for retransmitted packets, and thus minimize the possibility of delay caused by CAC algorithms. Queuing delay depends on available system resources. In other words, at a constant packet arrival rate, the more available bandwidth, the smaller blocking probability, and thus the smaller average delay. The rule considered in this work is first come, first served, and no priority is given to voice or data users. Optimal bandwidth allocation between voice and data users is not taken into account. Thus the average packet delay does not depend on which CAC algorithm is chosen, and hence it is not a proper figure of merit for comparison of different CAC algorithms.

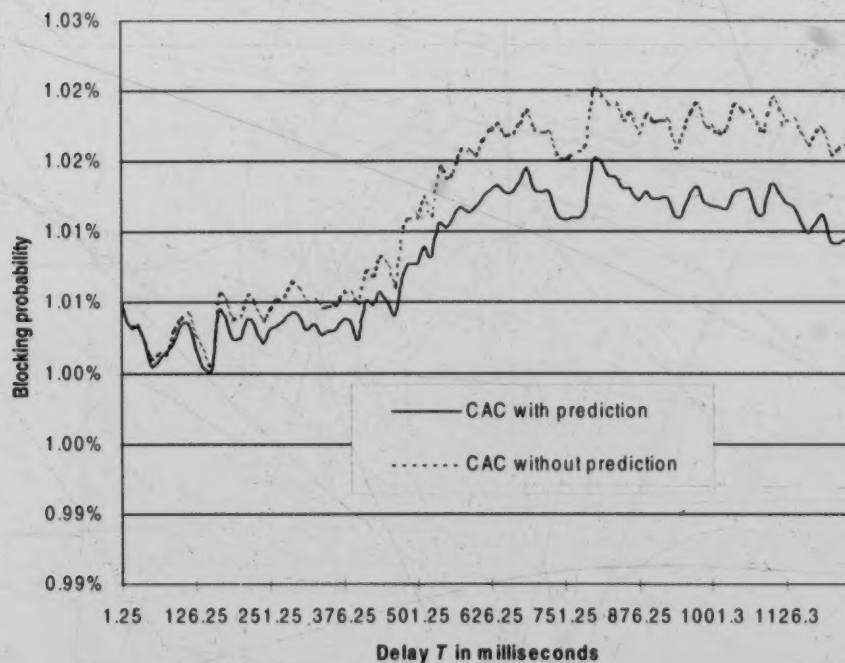


Fig. 4.9. Performance comparison of two call admission control schemes.

4.4 Conclusions

Capacity analysis of a multi-service CDMA system has been presented for both continuous and discontinuous transmission schemes. Discontinuous schemes can provide higher capacity for data traffic of very bursty nature depending on the overhead assigned to rapid synchronization of access request messages. For normal data traffic with an activity factor around 40%, discontinuous schemes have no advantage over continuous ones. The concept of equivalent bandwidth has been introduced which makes the analysis of an interference-limited system similar to that of a bandwidth-limited system without fixed channel assignment. In developing CAC policies for data users, the effect of the ARQ is considered which reduces the system's blocking probability, and eliminates the possibility of packet loss during transmission. From the simulations it is seen that if the call set up time is not very long, predictive CAC algorithm does not provide much advantage over non-predictive CAC scheme. If call holding time distribution and handoff rate are available, PCAC can reduce the blocking probability on the average compared with NPCAC if the call processing time is long.

References

- [1] D. Bertsekas and R. Gallager, *Data Networks*, Englewood Cliffs, NJ: Prentice Hall, 1987.
- [2] A. Leon-Garcia, *Probability and Random Processes for Electrical Engineering*, Addison-Wesley, Reading, Mass., 1994.

- [3] A. M. Viterbi and A. J. Viterbi, "Erlang capacity of a power controlled CDMA system", *IEEE Journal on Selected Areas in Communications*, Vol. 11, No. 6, August 1993, pp. 892-899.
- [4] K. S. Gilhousen, etc. , "On the capacity of a cellular CDMA system", *IEEE Trans on Vehicular Technology*, Vol. 40, No. 2, May 1991, pp. 303-312.
- [5] A. Sampath, N.B. Mandayam and J.M. Holtzman, "Erlang Capacity of a Power Controlled Integrated Voice and Data CDMA System", in *Proc. VTC'97*, May 1997, pp. 1557-1561.
- [6] S. Manji and Weihua Zhuang, "Capacity Analysis of an Integrated Voice/Data DS-CDMA Network", in *Proc. ICC'97*, June 1997, pp. 979-983.
- [7] Chih-Lin I and Richard D. Gitlin, "Multi-Code CDMA Wireless Personal Communications Networks", in *Proc. IEEE ICC'95*, June 1995, pp. 1060-1064.
- [8] T. S. Rappaport, *Wireless Communications: Principles & Practice*, Prentice Hall PTR, New Jersey, 1996.
- [9] P. T. Brady, "A statistical analysis of on-off patterns in 16 conversations", *Bell System Tech. Journal*, 47, 1968, pp. 73-91.
- [10] A. J. Viterbi, *CDMA: Principles of Spread Spectrum Communication*, Addison-Wesley, 1995.
- [11] TIA/EIA/IS-95, "Mobile Station-Base Station Compatibility Standard for Dual-Mode Wideband Spread Spectrum Cellular System.", *Telecommunication Industry Association*, July 1993.

- [12] A. Jalali and P. Mermelstein, "Effects of diversity, power control, and bandwidth on the capacity of microcellular CDMA wireless systems", *IEEE Journal on Selected Areas in Communications*, Vol. 12, No. 4, June 1994, pp. 952-961.
- [13] Songsong Sun, W. Krzymien, A. Jalali and Qiang Shen, "Performance Comparison of Two Medium Access Control Schemes for Packet Data Traffic in DS-CDMA Personal Communications Systems", in *Proc. Globecom'97/CTMC'97*, November 1997, pp. 604-608.
- [14] Songsong Sun, W. Krzymien, R. Krenz and A. Jalali, "High hybrid ARQ data transmission with optimum signal-to-interference ratio assignment in CDMA Personal Communications Systems", in *Proc. IEEE ICUPC'98*, Florence, Italy, October 1998, pp. 599-604.
- [15] Z. Liu and M. El Zarki, "SIR-based call admission control for DS-CDMA cellular systems", *IEEE Journal on Selected Areas in Communications*, Vol. 12, No. 4, May 1994, pp. 638-644.
- [16] M. Naghshineh and M. Schwartz, "Distributed call admission control in mobile/wireless networks", *IEEE Journal on Selected Areas in Communications*, Vol. 14, No. 4, May 1996, pp. 711-717.
- [17] A. Sutivong and J. M. Peha, "Performance comparisons of call admission control algorithms in cellular systems", in *Proc. GlobeCom'97*, November 1997, pp. 1645-1649.
- [18] TR45.5.0.18, "The CDMA2000 ITU-R RTT Candidate Submission", June 1998.
- [19] Z. Dziong, M. Jia and P. Mermelstein, "Adaptive traffic admission for integrated services in CDMA wireless-access networks", *IEEE Journal on Selected Areas in Communications*, Vol. 14, No. 9, December 1996, pp. 1737-1747.

Chapter 5

OPTIMAL FORWARD LINK POWER ALLOCATION

5.1 Introduction

An efficient utilization of system resources, such as power, bandwidth, and Walsh codes can increase the system capacity and improve the QoS. Some CDMA system resource allocation issues have already been addressed in the literature. Optimal SIR assignment for data transmission on the reverse link with the objective of maximizing system capacity is studied in Chapter 4. The allocation of transmitted powers and processing gains for different classes of users on the reverse link with perfect power control is considered in [1]. One of the key issues on the reverse link of CDMA systems is the implementation of an efficient power control algorithm (power management). The CDMA system capacity on the reverse link is mainly determined by the minimum required SIR [2]-[4]. In general, the CDMA system capacity is interference limited. However, on the forward link, it may be limited by the available number of Walsh codes, or available transmitted power. For example, in the IS-95[5], maximum 61 users at basic data rate can be serviced simultaneously (see Chapter 4) within each cell or sector. As far as power is concerned, the total available power from a base station transmitter is limited. If most data users are downloading large data files which means they need more power from the base station, and they are in remote locations as shown in Fig. 5.1, then the cell may support only a few users (not enough power available). On the other hand, if these users are located close to the base station, which means for the same size of a file they need much less power, the system may support more data users.

With the expanding usage of Internet, which normally requires downloading of large data files, efficient power allocation on the forward link under the constraint of limited base station total transmitted power is becoming more and more important.

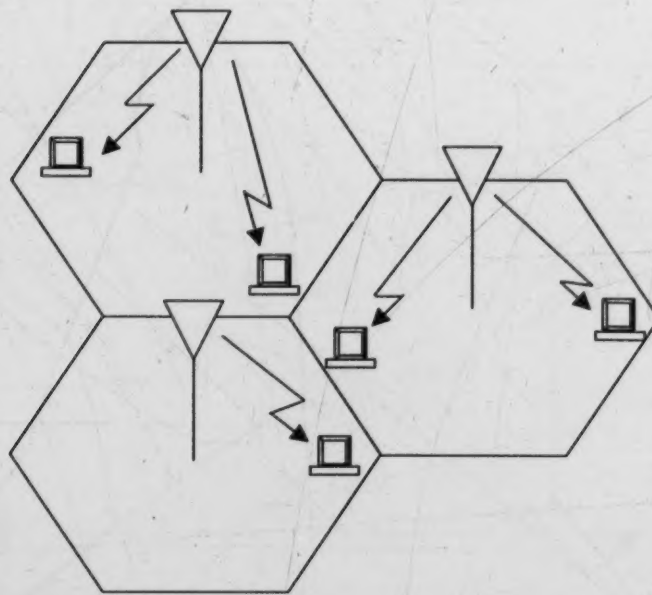


Fig. 5.1. Illustration of power utilization of data users.

For voice only CDMA systems, retransmission is not possible because of the critical signal delay limitation. Different users of a given cell are allocated different power levels according to their relative locations, and also their required quality of service (QoS), which can be mapped into the required SIR. The mobile station at the cell boundary requires a higher fraction of the transmitted power because of its larger distance from the base station and also, since it receives more interference from other base stations. In other words, for voice only CDMA systems, power is allocated to a given user to maintain

certain required SIR, and more power is provided to users receiving the most interference. Hence, in general, there does not exist a problem of optimal power allocation for voice communications. For data transmission, a hybrid ARQ scheme [6] can be employed to meet the strict QoS requirements. When retransmissions are allowed, the effective average power allocated to a data user needs to be calculated as the actual link transmitted power times the average number of transmissions of a data packet before it is accepted. If the actual link power allocated to a data user is high which naturally leads to a high SIR, the number of retransmissions is reduced. On the other hand, if that power is smaller, then the average number of retransmissions increases. Clearly, there is an optimum actual power that can be allocated to every data user in the cell, which minimizes the total average power transmitted by the base station. The reduction of the fraction of the total forward link power that needs to be allocated to an individual user increases the system capacity. This chapter addresses the issue of finding the optimal allocation to individual user of the power transmitted by the base station, which results in the maximum system capacity.

The remainder of this chapter is organized as follows. The forward link system description is presented in Section 5.2. Optimal power allocation is considered in Section 5.3. Conclusions and discussion are given in Section 5.4.

5.2 Forward link system description

The modem is similar to that described in [7], with an exception that an error detection encoder is added before the convolutional encoder in the transmitter, and error detection

takes place after the Viterbi decoder in the receiver. The block diagram of the data transmission is shown in Fig. 5.2. All users in the system employ BPSK modulation with pilot-signal-assisted coherent reception. Note that the pilot signal is unmodulated, it is covered by Walsh code 0 (see Fig. 5.3), while the traffic signals are coded, and spread by Walsh and PN codes (see Fig. 5.4). In Fig. 5.4, i is not zero, it could be range from 1 to 63 in the IS-95 system [5]. The pilot signal is shared by all traffic channels. Negative acknowledgment hybrid ARQ protocol is applied to meet the very low probability of accepted erroneous packet requirement (see Chapter 3).

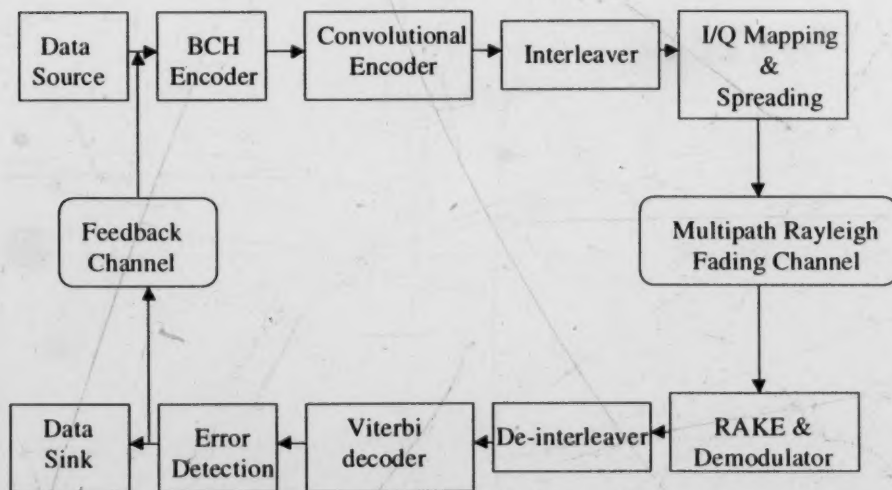


Fig. 5.2. Block diagram of the forward link CDMA data transmission system.

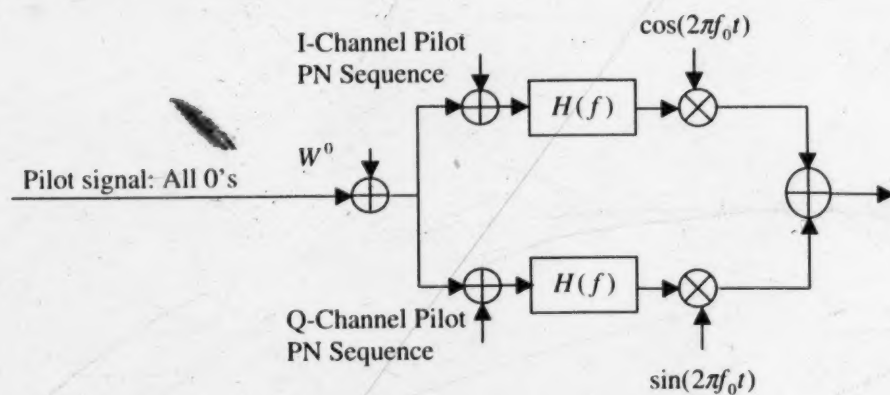


Fig. 5.3. Modulation and spreading for the pilot signal.

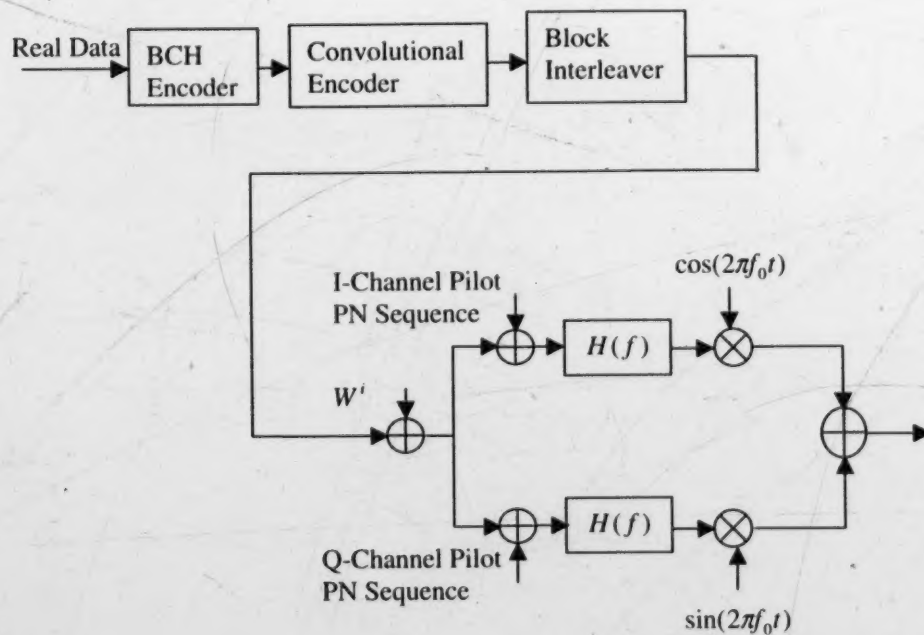


Fig. 5.4. Coding, modulation and spreading for the traffic signal.

Experimental data demonstrates that the received power S_R by the mobile at a distance r from a particular base station is proportional to [8]:

$$S_R \propto \frac{S_T 10^{\zeta/10}}{r^4} \quad (5.1)$$

where S_T is the transmitted power from a particular base station, and the term $10^{\zeta/10}$ represents signal power attenuation due to lognormal shadowing [8], ζ is a Gaussian random variable with zero mean and standard derivation σ (typically, $\sigma = 8$ dB). (5.1) is equivalent to (3.21) of Chapter 3 with path loss exponent equal to 4, and the path loss exponent may be larger than 4 in some environments. We assume that the given mobile i is connected to base station 0, and there are L base stations surrounding base station 0. Let the total transmitted power from base station j be S_{T_j} , and the given mobile i be at a distance r_{ji} from base station j . It is also assumed that the pilot signal is assigned a fraction $1 - \beta$ of the total base station transmitted power. The remaining fraction β is then allocated to all mobiles in the cell. Mobile i is allocated a fraction $\beta\varphi_i$ of the transmitted power, where $\varphi_i < 1$. All signals transmitted by a given base station share the same frequency band and propagate through the same multipath channel before arriving at the given mobile. Thus, the received SIR for mobile i in cell 0, denoted by ϵ_i , can be approximated as [2], [3]:

$$\epsilon_i = \frac{G\beta\varphi_i S_{T_0}}{S_{T_0}(1 - \beta\varphi_i) + 10^{-\zeta_0/10} \sum_{j=1}^L \left(\frac{r_{0i}}{r_{ji}} \right)^4 S_{T_j} 10^{\zeta_j/10}} \quad (5.2)$$

where G is the CDMA system processing gain, and ζ_j for $j=0, 1, \dots, L$, are mutually independent Gaussian random variables¹ with the same mean and variance as ζ in (5.1); they determine lognormal shadowing of interference from L interfering base stations to mobile i in cell 0. The background noise has been neglected in (5.2), which is reasonable since its value is trivial compared to the interference power received from all base stations. The powers S_{T_j} ($j=1, 2, \dots, L$) transmitted from all other L base stations, and the power $(1 - \beta\varphi_i)S_{T_0}$ (transmitted in cell 0 as pilot and traffic signals to other users) cause interference. Note that although the intra-cell interference would not exist under flat fading due to the orthogonality of Walsh functions for channelization on the downlink of the IS-95 system [5], it exists and is approximately equal to $(1 - \beta\varphi_i)S_{T_0}$ due to inter-path interference between resolvable paths under frequency-selective fading.

For voice communications, the signal power transmitted by the home base station to user i should be kept at a level sufficient to ensure that the average ϵ_i is adequate to obtain acceptable frame error rate ($\text{FER} \approx 10^{-2}$). The required average ϵ_i is normally around 5 dB [2], [3], [7]. In the case of data transmission, the average number of transmissions per successfully delivered packet is $N(\epsilon_i)$ for mobile i , and it is a function of ϵ_i . The issue for data transmission is how to allocate $\beta\varphi_i S_{T_0}$ to a particular user so that the effective average transmitted power $N(\epsilon_i)\beta\varphi_i S_{T_0}$ to user i by the base station is minimized.

¹ This implicitly assumes that the signals from different base stations, arriving at a given mobile, fade independently.

Minimization of the effective average power transmitted by the base station per user results in maximization of the number of users (capacity) that a base station can serve at a given level of its total power.

ϵ_i in (5.2) is a random variable, and hence, also $N(\epsilon_i)$ is random. Thus, the optimal forward link power allocation can be described as the stochastic nonlinear programming optimization problem, which will be discussed in detail in the next section.

5.3 Optimal power allocation

5.3.1 Problem formulation

As stated in Section 5.2, the optimal allocation of power is carried out under the criterion of minimizing the effective average transmitted power $N(\epsilon_i)\beta\varphi_i S_{T_n}$, which is equivalent to minimizing the normalized effective power $N(\epsilon_i)\varphi_i$. In general, the average number of transmissions per packet for data transmission employing an ARQ protocol can be expressed as:

$$N(\epsilon_i) = \sum_{i=1}^{\infty} i P_r^{i-1} (1 - P_r) = \frac{1}{1 - P_r} \quad (5.3)$$

where P_r is the retransmission rate, and it depends on the modulation and channel coding schemes. Unfortunately, a closed form of P_r as a function of ϵ_i can not be determined if convolutional codes are employed for channel coding. We may, however, approximate the retransmission rate as a function of the SIR in the following form [1]:

$$P_r \approx a \exp(-b\epsilon_i) \quad (5.4)$$

where a and b are constants which depend on modulation and coding.

If we assume equal transmitted powers from all base stations ($S_{T_0} = S_{T_1} = \dots = S_{T_L}$), which implies that all cells are loaded with equal traffic, (5.2) can be simplified to yield:

$$\varepsilon_i = \frac{G\beta\varphi_i}{1 - \beta\varphi_i + \sum_{j=1}^L \left(\frac{r_{0i}}{r_{ji}} \right)^4 10^{z_j/10}} \quad (5.5)$$

where $\xi_j = \zeta_j - \zeta_0$, $j=1, 2, \dots, L$, and hence they are also normally distributed with zero mean and standard deviation of $\sqrt{2}\sigma$.

The optimal power allocation is then a problem of finding φ_i which minimizes the normalized effective power $N(\varepsilon_i)\varphi_i$, which is a function of $\varphi_i, \xi_1, \xi_2, \dots, \xi_L$:

$$N(\varepsilon_i)\varphi_i = S_o(\varphi_i, \xi_1, \xi_2, \dots, \xi_L) \quad (5.6)$$

Using the approximation (5.4) we obtain $N(\varepsilon_i)\varphi_i = \frac{\varphi_i}{1 - ae^{-b\varepsilon_i}}$. This is a stochastic nonlinear programming problem, since $S_o(\varphi_i, \xi_1, \xi_2, \dots, \xi_L)$ is a random variable. The approach used in stochastic optimization is to convert the stochastic problem into an equivalent deterministic problem. In general, as described in [9], the new deterministic function for minimization, corresponding to (5.6) can be constructed as:

$$F(\varphi_i) = k_1 E[S_o(\varphi_i, \xi_1, \xi_2, \dots, \xi_L)] + k_2 \sqrt{\text{Var}[S_o(\varphi_i, \xi_1, \xi_2, \dots, \xi_L)]} \quad (5.7)$$

where k_1 and k_2 are nonnegative constants, and their values indicate the relative importance of mean value and standard deviation of $S_o(\varphi_i, \xi_1, \xi_2, \dots, \xi_L)$ for

minimization. For example, if we are interested only in minimizing the average transmitted power, we may let $k_1 = 1$, and $k_2 = 0$, and if the minimization of the expected value and standard deviation of $S_o(\varphi_i, \xi_1, \xi_2, \dots, \xi_L)$ is of equal importance, we may set $k_1 = k_2 = 1$.

To simplify the analysis, $S_o(\varphi_i, \xi_1, \xi_2, \dots, \xi_L)$ in (5.6) can be approximated as [9]:

$$S_o(\varphi_i, \xi_1, \xi_2, \dots, \xi_L) \approx S_o(\varphi_i, \bar{\xi}_1, \bar{\xi}_2, \dots, \bar{\xi}_L) + \sum_{j=1}^L \left(\frac{\partial S_o}{\partial \xi_j} \bigg|_{\bar{\xi}_1, \bar{\xi}_2, \dots, \bar{\xi}_L} \right) (\xi_j - \bar{\xi}_j) \quad (5.8)$$

where $\bar{\xi}_j$ is the mean value of ξ_j . The mean of $S_o(\varphi_i, \xi_1, \xi_2, \dots, \xi_L)$ is thus approximated as:

$$E[S_o(\varphi_i, \xi_1, \xi_2, \dots, \xi_L)] \approx S_o(\varphi_i, \bar{\xi}_1, \bar{\xi}_2, \dots, \bar{\xi}_L) \quad (5.9)$$

Since $\xi_1, \xi_2, \dots, \xi_L$ are mutually independent normal variables, the variance of $S_o(\varphi_i, \xi_1, \xi_2, \dots, \xi_L)$ is given by:

$$\text{Var}[S_o(\varphi_i, \xi_1, \xi_2, \dots, \xi_L)] = \sum_{j=1}^L \left(\frac{\partial S_o}{\partial \xi_j} \bigg|_{\bar{\xi}_1, \bar{\xi}_2, \dots, \bar{\xi}_L} \right)^2 \sigma_{\xi_j}^2 \quad (5.10)$$

Substituting (5.9) and (5.10) for the mean and standard deviation of $S_o(\varphi_i, \xi_1, \xi_2, \dots, \xi_L)$ in (5.7), we see that $F(\varphi_i)$ contains only one variable φ_i for which the optimal solution can be obtained using a rather simple graphical method explained in the following.

5.3.2 Numerical results

The CDMA system model described in Section 2 is considered. The system contains 19 cells as shown in Fig. 5.5, in which BS stands for a base station. It is assumed that the

interference to mobiles in cell 0 from base stations outside the second ring of cells has little effect on performance. We consider mobiles located at A, B, and C in cell 0. Before beginning minimization of $F(\varphi_i)$ given by (5.7), we need to find the suitable constant values of a and b in order to approximate the probability of retransmission according to (5.4). As mentioned earlier, a and b depend on the specific modulation and coding schemes. As stated in Section 5.2, the modulation and channel coding schemes are similar to those described in [7] with the exception that the error detection mechanism is added in the transceiver. To be specific, the user data is first encoded for error detection with a shortened (184, 160) binary BCH code², and then encoded for error correction with a rate 1/2 convolutional code of constraint length 9, followed by a bit interleaver. The interleaved encoded bits are spread, and then transmitted using BPSK. At the receiving end, after deinterleaving and decoding in the Viterbi decoder, the resulting bits are fed into the error detector. If errors are detected in the decoded frame, a retransmission is requested. The simulation results in [3] and [7] suggest that if we use the constant values of a and b as given in Table 5.1, the retransmission rate is well approximated by (5.4). All parameters used to obtain numerical results are listed in Table 5.1. The value of β used follows the assumptions of [2],[3] and [7], where 20% of the total power is devoted to the pilot channel.

In principle, we are interested in minimizing the effective average transmitted power from the base station to a particular mobile. Hence, $k_1 = 1$, and $k_2 = 0$ are used in the objective function (5.7). The normalized effective average transmitted powers

² The performance of this code is discussed in Chapter 3.

$F(\varphi_i)(=E[S_0(\varphi_i, \xi_1, \xi_2, \dots, \xi_L)])$ versus the normalized link transmitted powers φ_i are plotted in Figs. 5.6, 5.7 and 5.8, respectively for mobiles located at A, B, and C in cell 0.

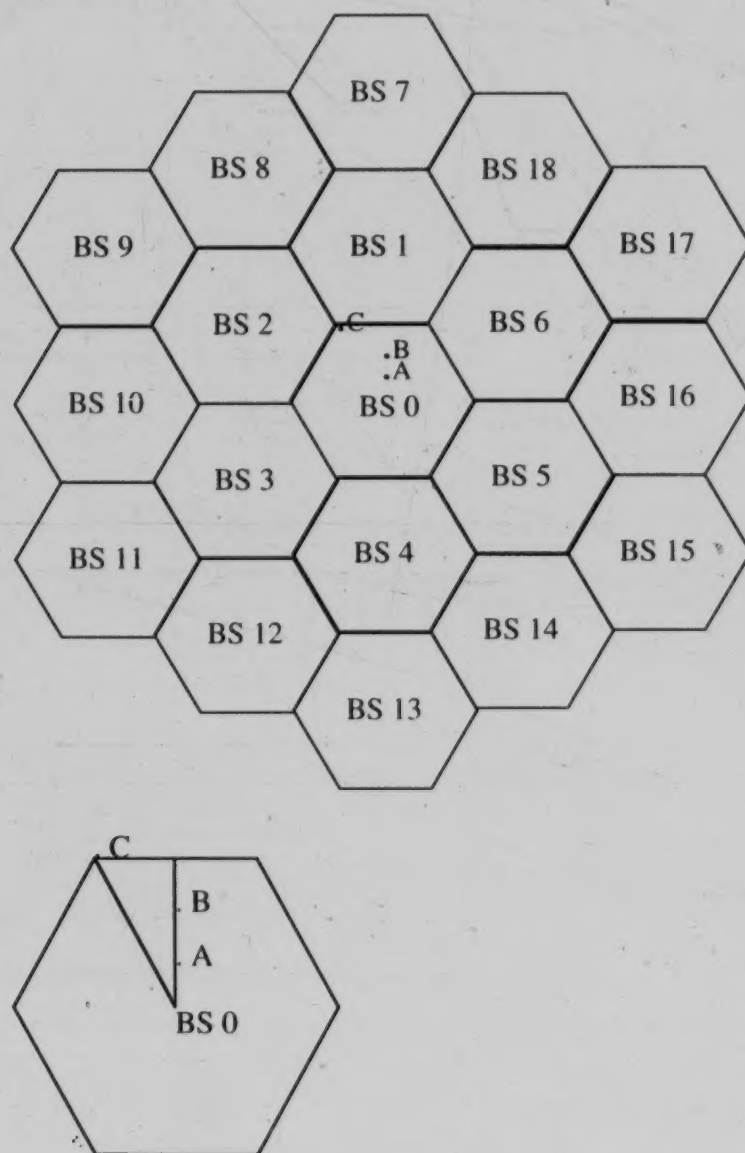


Fig. 5.5. The cell layout.

Table 5.1. Parameters used to obtain numerical results for the optimal power allocation

Parameter	Value
G	128
β	80%
a	99.6112
b	2.91133
σ	8 dB

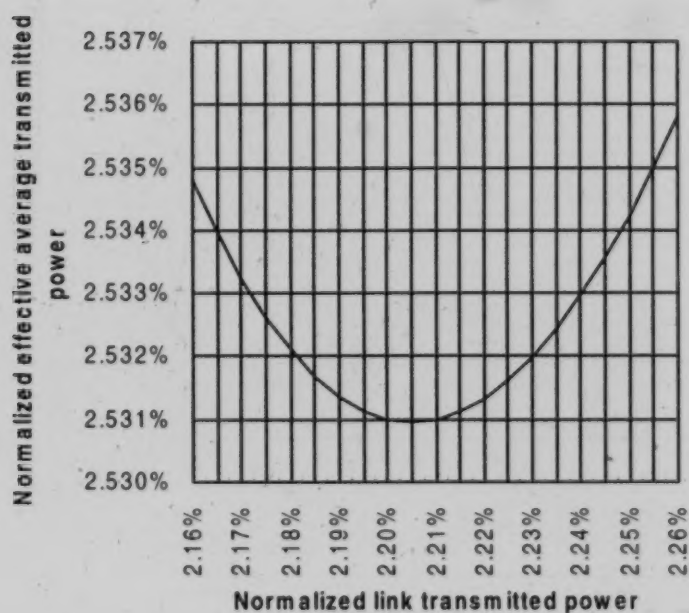


Fig. 5.6. Normalized effective average versus normalized link transmitted power for the mobile at location A.

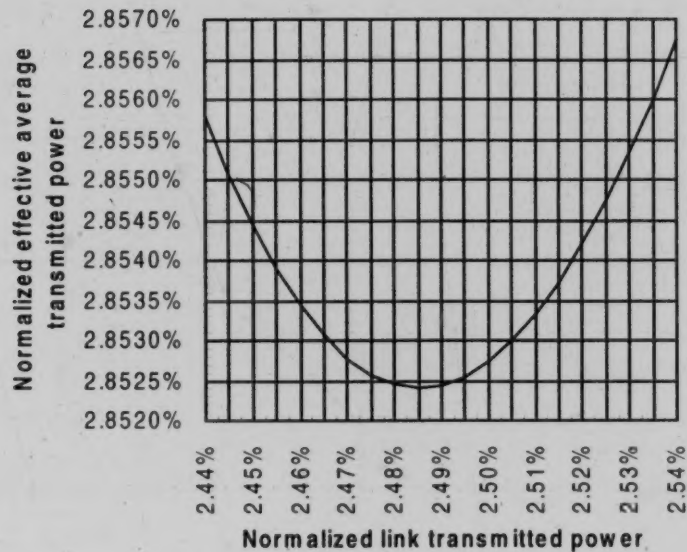


Fig. 5.7. Normalized effective average versus normalized link transmitted power for the mobile at location B.

Figs. 5.6, 5.7, and 5.8 indicate clearly that there is an optimal link transmitted power which minimizes the average transmitted power from the base station to a given mobile, and the required minimum transmitted power is location dependent. For example, at location A as shown in Fig. 5.6, the minimum effective average transmitted power is about 2.5309% of the total and it occurs when φ is around 2.205%. This also means the system can support 39.53 (reciprocal of 2.5309%) equivalent data users of the same class (same QoS requirement) if all mobiles are located in the ring of radius A surrounding the base station 0. At location B, optimal link power is about 2.485%, and the minimum effective transmitted power is about 2.8524%, which means the system can support about

35 data users if all of them are at a distance B from the base. At location C, the worst case, because it is equidistant from three base stations, the mobile needs around 8.4106% of the total power from the base station. Actually, mobiles at that position are in the state of soft handoff. The second base station must transmit to mobiles in the need of handoff the same power as the home base station.

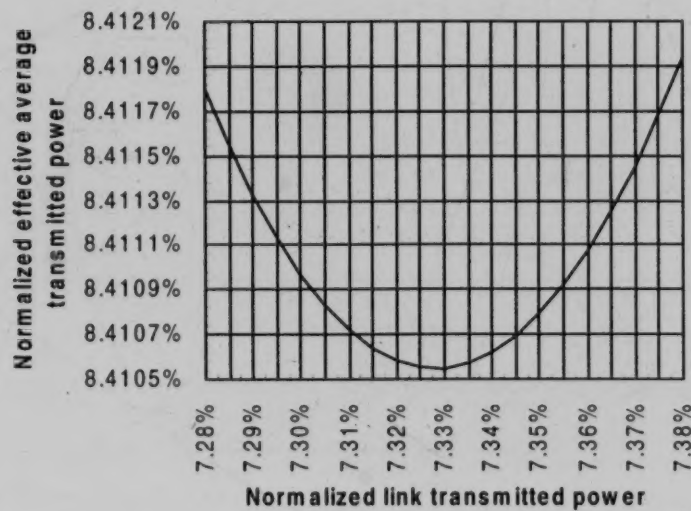


Fig. 5.8. Normalized effective versus normalized link transmitted power for the mobile at location C.

Although the minimum required effective transmitted power is location dependent, we find that the mean SIR, ϵ_i is almost the same at different locations when the effective transmitted power is minimized. For example, the optimal link transmitted powers for mobiles at location A, B, and C, as shown in Fig. 5.6, Fig. 5.7, and Fig. 5.8, are 2.205%, 2.485% and 7.33%, respectively. The mean required SIRs at these locations are all

around 3.6 dB. After checking several locations within the 30° right angled triangle³ shown in Fig. 5.5, it is found that the link SIRs are very close when the effective transmitted powers are minimized at these locations. This property could be very useful for forward link power control. Since the average received SIR is not location dependent when optimal power is allocated to a mobile, this suggests that the power control mechanism needs only to ensure that the average received SIR is at a value corresponding to the optimal power allocation.

If we are interested in minimizing both the average effective transmitted power and its standard deviation and treat the mean and standard deviation of the effective power as equally important, then $k_1 = k_2 = 1$ is used in the objective function (5.7). The optimal link transmitted powers, the effective mean powers, and the corresponding required SIRs corresponding to this case are listed in Table 5.3. Note that this time, the average effective power is not necessarily at the minimum. Also, the link SIR depends now on the location when the optimal power is allocated. For comparison, the results corresponding to the case of minimizing only the mean transmitted power are also presented in Table 5.2. We see that, in order to minimize the mean and the standard deviation of the total transmitted power, the actual link power increases, and so does the average effective transmitted power. The average required SIR increases with the increasing distance from the base station. This makes the forward link power control more difficult.

³ By symmetry, it is only necessary to check over the area of this triangle.

Table 5.2. Numerical results when the criterion is to minimize the mean transmit power

Location	Optimal power (normalized)	Total average power (normalized)	Link SIR in dB
A	2.205%	2.53%	3.59
B	2.485%	2.85%	3.59
C	7.33%	8.4%	3.60

Table 5.3. Numerical results when the criterion is to minimize the sum of the mean transmit power and its standard deviation

Location	Optimal power (normalized)	Total average power (normalized)	Link SIR in dB
A	2.22%	2.54%	3.62
B	2.76%	2.94%	4.05
C	9.9%	10.1%	4.92

5.4 Conclusion and discussion

Due to the necessity of retransmissions to achieve highly reliable data transmission in CDMA systems, there is an optimal power allocation which minimizes the total transmitted power to a particular mobile from the base station. Optimal forward link power allocation will lead to increased system capacity, since the total power available to all mobiles from a base station is limited. The optimal power is allocated to mobiles

depending on their positions in the cell, and more power must be provided to mobiles receiving the most interference from all base stations. If the optimization criterion is to minimize the mean transmitted power, then the link SIR corresponding to the optimal power is approximately location independent. This suggests that the forward link power control can be implemented by keeping the mobile's received SIR constant, and making sure corresponding to this constant SIR, the mean transmitted power is minimized. On the other hand, if the criterion is to minimize both the mean and the standard deviation of the transmitted power, both the optimal power and the link SIR are location dependent.

Optimal power allocation determined in this chapter maximizes the system capacity. However, there is a tradeoff between maximizing capacity and reducing the retransmission rate. Minimizing the mean transmitted power increases the system capacity, but the retransmission rate may increase. For example in Fig. 5.6, the optimal power is around 2.205%, and for this value the retransmission rate is around 0.1. However when $\varphi = 2.26\%$, the retransmission rate is about 0.08, and the average effective transmitted power is also increased. High retransmission rate increases the system overhead, such as more buffer capacity, the number of negative acknowledgment packets, etc. A practical compromise is to find the suitable transmitted power φ (not necessarily optimal) in a figure like Fig. 5.6 which makes both the mean transmitted power and retransmission rate acceptable.

References

- [1] M.L. Honig and J. B. Kim, "Resource Allocation for Packet Data Transmission in DS-CDMA", in *Proc. 1995 Allerton Conference on Communications, Control, and Computing*, Monticello, IL, October 1995, pp. 925-934.
- [2] K. S. Gilhousen, *et al.*, "On the capacity of a cellular CDMA system", *IEEE Trans. on Vehicular Technology* 40(2), pp. 303-312.
- [3] A. J. Viterbi, *CDMA: Principles of Spread Spectrum Communication*, Addison-Wesley, 1995.
- [4] Songsong Sun, W. Krzymien, "Call admission policies and capacity analysis of a multi-service CDMA PCS with continuous and discontinuous transmission", in *Proc. IEEE VTC'98*, Ottawa, Canada, May 1998, pp. 218-223.
- [5] TIA/EIA/IS-95, "Mobile Station-Base Station Compatibility Standard for Dual-Mode Wideband Spread Spectrum Cellular System," *Telecommunication Industry Association*, July 1993.
- [6] S. Lin and D. J. Costello, Jr. *Error Control Coding: Fundamental and Applications*, Prentice-Hall, New Jersey, 1983.
- [7] A. Jalali and P. Mermelstein, "Effects of Diversity, Power Control, and Bandwidth on the Capacity of Microcellular CDMA Systems", *IEEE Journal on Selected Areas in Communications*, Vol. 12, No. 5, June 1994, pp. 952-961.
- [8] W. C. Y. Lee, *Mobile Cellular Telecommunications Systems*, McGraw-Hill, New York, 1989.
- [9] S. S. Rao, *Optimization: Theory and Applications*, Wiley Eastern Limited, New Delhi, 1979.

Chapter 6

CONCLUSIONS AND FUTURE RESEARCH DIRECTIONS

This thesis research project has addressed issues of high data rate transmission in CDMA wireless communication systems. The study involves the development of efficient access systems, reliable data transmission schemes, call admission control policies and optimal power allocation. Three of the aforementioned topics arise in the context of the reverse link transmission, and the last one is related to the forward link. Since this project is partially sponsored by industry, the research focus is on using the commercially successful IS-95 CDMA system which has been developed and deployed in Asia and North America, to transmit data. Although a fairly extensive study has been conducted on the research questions covered by the thesis, several system design issues still remain to be studied. These are discussed below.

The IS-95 approach works very well for voice services and is claimed to be a near optimal solution for voice communications [1], [2], but not necessarily for data transmission. In other words, some kind of modification of the existing IS-95 system is needed in order to improve the system performance for data transmission. This issue has been addressed by several third generation CDMA standard proposals [2], [3]. One of the changes is that on the reverse link, the pilot-aided coherent detection technique will be used in third generation systems while the existing IS-95 system uses 64-ary orthogonal modulation. The reason for this is straightforward: the modulation scheme using coherent

detection requires lower SIR than with noncoherent detection. As discussed in [4], the required SIR is about 1 dB less than for the 64-ary noncoherent modulation used in IS-95.

Another important issue is related to channel coding. For data transmission without critical latency requirements, convolutional codes with Viterbi decoding are not the best choice for channel coding. Near Shannon limit error-correcting codes, such as turbo-codes [5]-[8], and serial concatenated codes [9] could be strong candidates. These codes are claimed to achieve near Shannon limit performance with relatively simple component codes and long interleavers. A required bit energy-to-noise ratio of 0.7 dB was reported for BER of 10^{-5} in [5] and [6], which is 0.5 dB above Shannon limit. With such low required signal-to-noise ratio, a CDMA system capacity can be significantly increased. When turbo codes are used for channel coding, the hybrid ARQ scheme using some novel techniques [10] which involve some form of packet combining, is more efficient than the type I ARQ. This can be explained as follows. The decoding algorithm for turbo codes is iterative and the useful information from the previous decoder is fed to the latter one to enhance the performance of the decoder. Hence, the decoded soft information from prior transmission and the retransmission can be effectively combined at the receiver [10], and the whole system performance can be greatly improved.

When turbo codes are employed for data transmission, one question naturally arises: how many iterations are required for decoding per transmission? For example, for a rate $\frac{1}{2}$ turbo code, memory length of 4, generators $G_1 = 37$ (octal), $G_2 = 21$ (octal), with interleaving 256×256 , for one time transmission, a signal-to-noise (SNR) of 0.7 dB is

required to obtain the BER of 10^{-5} when the number of iterations is 18, but the required SNR is about 1.05 dB when the number of iterations is 6 [5], [6]. Which one is better for data transmission (6 or 18 iterations?), since neither can meet the strict data transmission requirement such as BER below 10^{-10} ? A study on the hybrid ARQ scheme using turbo codes, considering number of iterations, acknowledgement overhead and optimization of SNR will be very interesting.

To implement the CAC algorithm developed in Chapter 4, the home base station has to know the equivalent bandwidth introduced by a user from neighbouring or remote cells ($\varphi_i W_d$ in Chapter 4). The key issue is how to obtain the coefficient φ_i for each particular user. The mean value for φ_i used in Chapter 4 is obtained by assuming equal traffic loading of the whole network. For the system with unequal traffic loading, a method to estimate φ_i for real time implementation will improve the system performance.

The study is focused mainly on the reverse link. With the expanding usage of the Internet, requiring downloading of large data files, the forward link problem is becoming more and more important. Although forward link power allocation is addressed in this thesis, there are other issues which need to be considered, such as CAC algorithms on the forward link, fast closed loop power control algorithms, multicarrier or direct spread option, or even the billing system. The billing system for data users may be different from that for voice users which are normally charged depending on the connection time. For data users, the billing could be based on how many packets are received or transmitted, and user location. As presented in Chapter 5, when the mobile station is in the worst location,

it needs much more power than other mobiles. A mobile station close to the base station needs much less power than those far away from it. Hence, location may be a significant factor in the billing system for data users.

References

- [1] A. J. Viterbi, "Principles of Commercial CDMA", in *Proc. IEEE ICUPC'97 Tutorial*, October, 1997.
- [2] E. G. Tiedemann Jr., Yu-Cheun Jou, and J. P. Odenwalder, "The evolution of IS-95 to a third generation system and to the IMT-2000 era", *ACTS Mobile Communications Summit'97*, Aalborg, Denmark, October 1997, pp. 924-929.
- [3] T. Ojanperä, "Development of 3G radio technologies", *First Annual UCSD Conference on Wireless Communications*, San Diego, California, March 1998, pp. 127-132.
- [4] S. Sun, W. Krzymien, R. Krenz and A. Jalali, "Reliable hybrid ARQ data transmission with optimal signal-to-interference ratio assignment in CDMA PCS", in *Proc. IEEE ICUPC'98*, Florence, Italy, October 1998, pp. 599-604.
- [5] C. Berrou, A. Glavieux, and P. Thitimajshima, "Near Shannon limit error-correcting coding and decoding: Turbo-codes", in *Proc. IEEE ICC'93*, Geneva, 1993, pp. 1064-1070.
- [6] C. Berrou and A. Glavieux, "Near optimum error-correcting coding and decoding: Turbo-codes", *IEEE Trans. Communications*, vol. 44, no. 10, October 1996, pp. 1261-1271.

- [7] D. Divsalar and F. Pollara, "Turbo codes for PCS applications", in *Proc. IEEE ICC'95*, 1995, pp. 54-59.
- [8] P. Jung, "Comparison of turbo-code decoders applied to short frame transmission systems", *IEEE Journal on Selected Areas in Communications*, Vol. 14, No. 3, April 1996, pp. 530-537.
- [9] S. Benedetto and D. Divsalar, "Turbo-codes: Performance analysis, design and iterative decoding", in *Proc. IEEE Globecom'97 Tutorial*, November 1997.
- [10] K. R. Narayanan and G. L. Stuber, "A novel ARQ technique using the turbo coding principle", *IEEE Communications Letters*, Vol. 1, No. 2, March 1997, pp. 49-51.